DIAGNOSTIC ACCURACY ANALYSIS FOR ORDINAL COMPETING RISK OUTCOMES USING ROC SURFACE

by

Song Zhang

B. M. Medicine, Harbin University of Medicine, **China,** 1995

M. S. Biostatistics, University of Pittsburgh, 2012

Submitted to the Graduate Faculty of

the Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2017

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

UNIVERSITY OF PITTSBURGH

This dissertation was presented

by

Song Zhang

It was defended on

December 6, 2017

and approved by

Abdus Wahed, Ph.D

Professor and Director of Ph.D Program

Department of Biostatistics

Graduate School of Public Health

University of Pittsburgh

Yu Cheng, Ph.D

Associate Professor

Department of Statistics

The Dietrich School of Arts & Sciences

University of Pittsburgh

Joyce Chang, Ph.D

Professor

Department of Biostatistics

Graduate School of Public Health

University of Pittsburgh

Maria Mor, Ph.D

Assistant Professor

Department of Biostatistics

Graduate School of Public Health

University of Pittsburgh

Dissertation Advisor: **Abdus Wahed**, Ph.D

Professor and Director of Ph.D Program

Department of Biostatistics

Graduate School of Public Health

University of Pittsburgh,

Co-Advisor: **Yu Cheng**, Ph.D

Associate Professor

Department of Statistics

The Dietrich School of Arts & Sciences

University of Pittsburgh

DIAGNOSTIC ACCURACY ANALYSIS FOR ORDINAL COMPETING RISK OUTCOMES USING ROC SURFACE

Song Zhang, PhD

University of Pittsburgh, 2017

ABSTRACT

Many medical conditions are marked by a sequence of events or statuses that are associated with continuous changes in some biomarkers. However, few works have evaluated the overall accuracy of a biomarker in separating various competing events. Existing methods usually focus on a single cause and compare it with the event-free controls at each time. In our study, we extend the concept of ROC surface and the associated volume under the ROC surface (VUS) from multi-category outcomes to ordinal competing risks outcomes. We propose two methods to estimate the VUS. One views VUS as a numerical metric of correct classification probabilities representing the distributions of the diagnostic marker given the subjects who have experienced different cause-specific events. The other measures the concordance between the marker and the sequential competing outcomes. Since data are often subject to lost of follow up, inverse probability of censoring weight is introduced to handle the missing disease status due to independent censoring. Asymptotic results are derived using counting process techniques and U-statistics theory. Practical performances of the proposed estimators in finite samples are evaluated through simulation studies and the procedure of the methods are illustrated in two real data examples.

Public Health Significance: ROC curve has long been treated as a gold standard in evaluating the accuracy of continuous predictors in separating binary outcomes in various fields including biomedical, financial, and geographical areas. Our proposed methods extend its utilization in multi-category events outcomes to competing risks censoring. Our methods

aim to assess a global accuracy of a biomarker's predictive power to each simultaneously, especially, to which stages of disease progression that patients would land by a specific time in followup. Our work provides much-needed global assessment for the predictive power of a biomarker for disease progression.

Keywords: Concordance probability; Correct classification probability; Discriminative capability; Disease progression; Inverse probability of censoring weighting; U-statistics.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

In biomedical studies, it is often of interest to measure a biomarker's predictive power for future events. Accurate diagnostic analysis can help clinicians screen high-risk subjects, which in turn will lead to timely and efficient therapeutic interventions and reduced mortality and morbidity. A time-dependent ROC curve was proposed by Heagerty et al. (2000) to extend diagnostic accuracy analysis from a binary outcome to a typical survival outcome, summarizing sensitivity and specificity at a specific time t_0. Assuming a lower biomarker value is associated with a worse outcome, the area under the ROC curve (AUC) can be interpreted as a concordant probability that for a randomly selected pair of a case (i.e., developing the event of interest by t_0) and a control (no event by t_0), the biomarker value from the case is lower than that from the control. A lot of works have shown that the AUC is an expression of Mann-Whitney U statistics (Faraggi and Reiser, 2002). Heagerty and Zheng (2005) presented their work in constructing predictive accuracy measure for survival outcomes by fitting Cox regression models.

Competing risk censoring commonly arises in studies where subjects are at risk for multiple failures, and one failure precludes the observation of the others or alters the probability of occurrence of the others (Gooley et al., 1999). For instance, in the Monongahela-Youghiogheny Healthy Aging Team (MYHAT) study, if we took the onset of mild/moderate cognitive impairment (MCI) or dementia as a primary interest, when a patient died without cognitive decline, the onset of MCI/dementia would be competing-risk censored by death. Note that the mechanism of "redistribution to the right" (Efron, 1967), is limited to the assumption of independent censoring in Kaplan-Meier methods and Cox models, and thus is not valid in competing risks censoring. With independent censoring, the likelihood contribution from censoring data which amounts to a constant in estimating parameters of

interest, should not affect inference. However, the share of risks from competing events will not be distributed to the other subjects at the riskset due to interactions between the competing events. A fundamental proof from Tsiatis (1975) showed that the dependence among competing events was not non-parametrically identifiable given only observed data, so the observed data alone is not sufficient to identify the marginal distributions of the latent variables without imposing their dependence structure. More specifically, even under a parametric model with its maximum likelihood estimator producing consistent estimators, it is not possible to assess whether the models are correctly specified. A popular quantity for measuring the cumulative probability of the event of interest by a specific time is the cumulative incidence function (CIF) (Prentice et al., 1978; Kalbfleisch and Prentice, 2011), since the survival function for a specific event may not be well defined. It has been widely employed due to its intuitive probability interpretation and non-parametric identifiability.

Expanding the ROC concept to competing risks consoring, Saha and Heagerty (2010) proposed to estimate the sensitivity with cumulative cases accruing to a fixed time, and to estimate the specificity among "healthy" control subjects (i.e., those without developing any event yet). Following similar definitions of sensitivity and specificity, Zheng et al. (2012) evaluated prognostic accuracy with multiple covariates using both Cox (1959) and more flexible Scheike et al. (2008) models. Blanche et al. (2013a) and Wolbers et al. (2014) used nonparametric inverse probability of censoring weighting (IPCW) to derive the estimators of AUCs and their asymptotic properties. Some of these analyses compare cases from each cause to the event-free control at each time. One limitation of the above methods is that there is no overall predictive accuracy assessment across all of events simultaneously, since different cases are considered for each specific cause in separate ROC analyses. In contrast, Shi et al.(2014) evaluated the improved accuracy of new markers for competing outcomes by defining "controls" that combine both event-free subjects and those who have developed competing events. Though the definition of the "control" group is in line with the augmented "at-risk" set in Fine and Gray (1999), it may not be ideal if subjects with competing events are very different from those without any event. To the best of our knowledge, the existing methods are not adequate in evaluating a biomarker's discriminatory power on competing outcomes, since they either do not provide a global assessment of accuracy across all competing events, or have to force unnatural grouping of competing events with healthy controls.

Medical conditions often manifest a natural ordinal disease status. For example, in the MYHAT study, disease progresses through several sequential stages: normal cognitive function, mild to moderate cognitive impairment, severe dementia and death. Also, patients with cirrhosis follow a series of continuous progression marked by various stages of disease severity. Following onset of viral liver hepatitis, patients can be observed with declining liver functions to fibrosis, which is marked as normal liver cells replaced by functionless scar tissue in mild level (less than 25% scar tissue), and then deteriorating to cirrhosis which is characterized with severe liver cell damage with 50% to 75% scar tissue, and finally they die with complications due to collapse of liver function. An important aim in clinical practices is to characterize a sequence of progressive status based on continuous changes in a prognostic biomarker. With multi-level progressive status, Obuchowski (2005) pointed out that simply dichotomizing an ordinal outcome led to upward bias in diagnostic testing using ROC curve. Mossman (1999) introduced the concept of multi-dimensional ROC surface and the Volume Under the ROC Surface (VUS) by evaluating discriminatory accuracy of two diagnostic tests for three disease categories. The variance of Mossman (1999)'s VUS estimator was derived by Dreiseitl et al. (2000) based on the theory of U-statistics. Li and Fine (2008) applied the concept of VUS to unordered multilevel categorical outcomes and further expanded multi-way ROC analysis and the summary statistics of Hypervolume Under the ROC Manifold (HUM). Li and Zhou (2009) studied the estimated VUS using nonparametric and semiparametric methods and developed asymptotic properties of the estimators. Wu and Chiang (2013) showed through a rigorous proof that HUM is directly related to an explicit U-estimator.

However, there is no well-developed method to assess the global accuracy in a biomarker's prediction of which stage of disease progression a subject would land by a specific time. Thus, we propose to utilize the concept of the ROC surface (or ROC manifold) and the VUS (or HUM) to demonstrate the prognostic accuracy of a biomarker to competing risks outcomes. Here, we focus on the competing events with a natural order, as we have observed in the MYHAT study. The rest of the dissertation is organized as follows. In §2, we introduce the concept of the ROC surface and the associated VUS from two perspectives. One is to employ

the building blocks of an ROC surface, namely correct classification probabilities (CCPs), to derive VUS. The second is to measure a concordance probability between the biomarker and the competing outcomes. Subsequently, two methods are proposed to estimate the VUS from these two perspectives in §3 and §4, and their asymptotic properties are investigated. A real data analysis phenomenon of adaption tied scores in biomarker using our proposed methods is discussed in §5. In §6, we study the finite sample performances of our proposed estimators through simulations, and illustrate our methods through the analyses of two real data examples in §7. In the end, we summarize our work and discuss some future research ideas in §8.

2.0 ROC SURFACE FOR ORDINAL COMPETING RISKS OUTCOMES

2.1 NOTATION

Without loss of generality, we now consider the diagnostic accuracy of one single biomarker in predicting two ordered competing risks outcomes, referred as a cause-1 event and a cause-2 event. The cause-1 event is assumed to be a worse medical condition, as compared to healthy controls. Let Y denote a diagnostic biomarker where lower values correspond to worse medical conditions. In practice, a biomarker may have tied values. Discussions on how to handle ties are given in §5 and the corresponding simulation results are given in §6. Let T be the time to any ordinal competing events, and $\epsilon = 1, 2$ be the corresponding cause of failure. At a fixed time t_0, if both competing events have occurred, the more severe progression time is recorded as T and ϵ is set to be 1. We then define disease status $D(t_0)$ as:

$$\begin{cases} D\left(t_0\right) = 1, & \text{if } T \leq t_0, \epsilon = 1, \\ D\left(t_0\right) = 2, & \text{if } T \leq t_0, \epsilon = 2, \\ D\left(t_0\right) = 0, & \text{if } T > t_0. \end{cases}$$

In practice, there may be administrative censoring C, in addition to competing risks censoring. Thus, we observe $X = \min\left(T, C\right)$ and the combined cause indicator $\eta = I\left(T \leq C\right)\epsilon$, where $I\left(\cdot\right)$ is an indicator function. The observed data consist of i.i.d replicates $\left\{\left(Y_i, X_i, \eta_i\right), i = 1, \ldots, n\right\}$.

2.2 ROC SURFACE

Analogous to sensitivity and specificity that describe the level of accuracy by specifying a series of cutpoints along with a continuous classifier for a binary outcome, we need two cutpoints $(c_1, c_2) \in \mathcal{R}^2$ from Y with $c_1 \leq c_2$ for the two competing events, where we assign a subject to Class 1 if their biomarker $Y \leq c_1$, to Class 2 if $c_1 < Y \leq c_2$, and to Class 3, otherwise. Correct classification probabilities (CCPs) are then defined for the subjects experiencing a cause-1 event, a cause-2 event, or none of the events at a given time t_0 as follows:

$$\begin{aligned} CCP_1 &= P\left(Y_i \leq c_1 | T_i \leq t_0, \epsilon_i = 1\right) = F_{Y|1}(c_1), \\ CCP_2 &= P\left(c_1 < Y_i \leq c_2 | T_i \leq t_0, \epsilon_i = 2\right) = F_{Y|2}(c_2) - F_{Y|2}(c_1), \\ CCP_3 &= P\left(Y_i > c_2 | T_i > t_0\right) = 1 - F_{Y|3}(c_2), \end{aligned}$$

where $F_{Y|d}(y) = P(Y \leq y | D(t_0) = d), d = 1, 2, 3$, is the conditional cumulative density function (CDF) of the biomarker Y, given that the subject is in a particular disease status. Also, similar to the ROC curve characterizing the full spectrum of sensitivity and specificity in a two-dimensional space, the plot of (CCP_1, CCP_2, CCP_3) in a three-dimensional space at all possible values of (c_1, c_2) generates a three-dimensional ROC surface for three-category time-dependent outcomes. Following Li and Zhou (2009), the ROC surface is defined by expressing CCP_2 as a function of CCP_1 and CCP_3:

$$Q(u, v) = \begin{cases} F_{Y|2}\{F^{-1}_{Y|3}(1-u)\} - F_{Y|2}\{F^{-1}_{Y|1}(v)\} & \text{if } F^{-1}_{Y|1}(v) \leq F^{-1}_{Y|3}(1-u), \\ 0 & \text{otherwise.} \end{cases} \tag{2.2.1}$$

The ROC surface can also be formulated by using CCP_1 or CCP_3 as a function of the other two CCPs. However, definition (2.2.1) is simpler than the other two. The VUS is then defined as:

$$VUS = \int_0^1 \int_0^1 Q(u, v) du dv. \tag{2.2.2}$$

To having a feel if the ROC surface and its associated VUS, we present the estimated ROC surfaces in Figures 2.1 and 2.2 under different different predictive powers of a biomarker to the ordinal competing outcomes, where cases of death (the cause-1 event) and dementia

(the cause-2 event) are compared to the healthy controls simultaneously. In Figures 2.1, left ROC surface represents strong predictive power of a biomarker to the ordinal competing events which VUS is equal to 0.79, and right ROC surface represents moderate predictive power between a biomarker and the competing events for VUS = 0.49. Also in Figure 2.2, left ROC surface represents weak predictive for VUS = 0.35 and right one represents no predictive power for VUS = 0.167. The details of estimating the ROC surface and VUS are discussed in §3. Figures seem to suggest that the VUS is a sensitive global mesasure of the strength of the association between the marker and the competing risks outcomes.

2.3 VUS CORRESPONDING TO A CONCORDANCE INDEX

According to previous works of Mossman (1999), Dreiseitl et al. (2000) and Wu and Chiang (2013), the VUS also corresponds to a concordance measure between the biomarker values and the sequential competing risks outcomes. Suppose we randomly select three subjects 1, 2, 3 such that $T_1 \leq t_0, \epsilon_1 = 1, T_2 \leq t_0, \epsilon_2 = 2$ and $T_3 > t_0$. Their biomarkers are denoted as Y_1, Y_2 and Y_3, respectively. It can be shown that

$$VUS = P(Y_1 < Y_2 < Y_3 | T_1 \leq t_0, \epsilon_1 = 1, T_2 \leq t_0, \epsilon_2 = 2, T_3 > t_0). \tag{2.3.1}$$

It is clear that the VUS is 1/6 for a completely non-informative test in relation to the three-category outcomes.

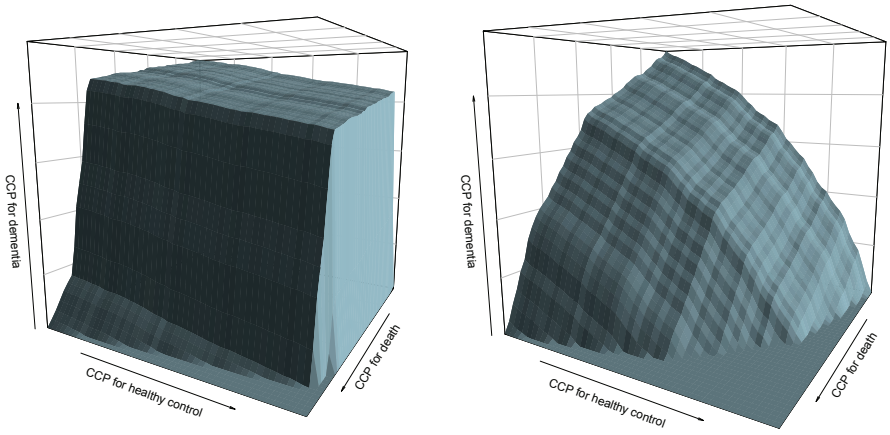

Figure 2.1: ROC surface for strong predictive power and moderate predictive power of a biomarker to ordinal competing events

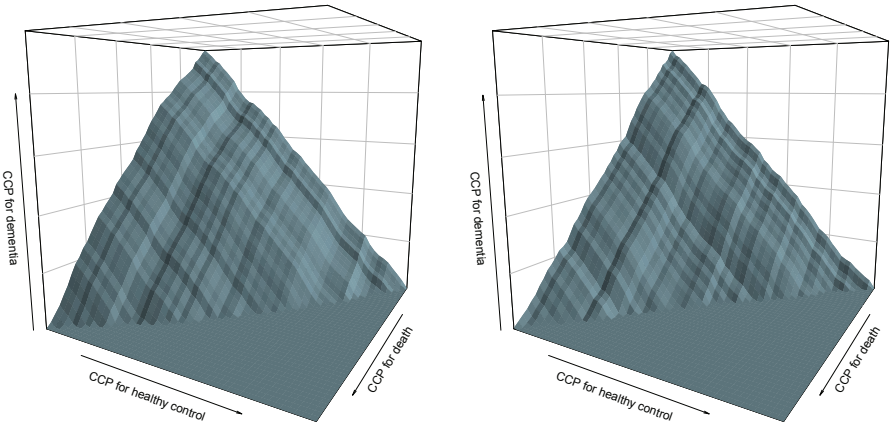

Figure 2.2: ROC surface for weak predictive power and no predictive power of a biomarker to ordinal competing events

3.0 ESTIMATION OF THE ROC SURFACE AND VUS

3.1 NONPARAMETRIC ESTIMATION OF THE THREE-DIMENSIONAL ROC SURFACE AND VUS

Nonparametric methods are used to estimate $F_{Y|d}(y), d = 1, 2, 3$ for any pair of (y, t_0). In terms of estimating the conditional distribution of Y given the subjects experiencing either the cause-1 event or the cause-2 event by t_0, we have

$$\begin{aligned} F_{Y|k}(y) &= P(Y \leq y | D(t_0) = k) \\ &= \frac{P(Y_i \leq y, T_i \leq t_0, \epsilon_i = k)}{P(T_i \leq t_0, \epsilon_i = k)}, \end{aligned}$$

where $k = 1, 2$. The numerator is a joint bivariate CIF and the denominator is a univariate CIF. We adopt the bivariate CIF estimator in Cheng et al. (2007), $\hat{F}_{Y,T}(t_0)$, to estimate the joint bivariate probability, which includes the biomarker Y as a special case being completely observed without censoring. The univariate CIF is estimated by the standard nonparametric estimator $\hat{F}_T(t_0)$. Also, we formulate the conditional distribution of Y for those subjects who have not developed any events by t_0 in terms of a bivariate survival function and a univariate survival function:

$$\begin{aligned} F_{Y|3}(y) &= P(Y \leq y | D(t_0) = 3) \\ &= \frac{P(Y_i \leq y, T_i > t_0)}{P(T_i > t_0)} \\ &= \frac{P(T_i > t_0) - P(Y_i > y, T_i > t_0)}{P(T_i > t_0)} \\ &= \frac{S_T(t_0) - S_{Y,T}(t_0)}{S_T(t_0)}. \end{aligned}$$

The bivariate survival estimator $\hat{S}_{Y,T}(t_0)$ (Dabrowska, 1988) is used to estimate the bivariate survival function $S_{Y,T}(t_0)$ and the Kaplan-Meier estimator $\hat{S}_T(t_0)$ is used for the univariate survival function $S_T(t_0)$. Plugging the estimators into the definition of the three-dimensional ROC surface, we have

$$\hat{Q}(u,v) = \hat{F}_{Y|2}\{\hat{F}^{-1}_{Y|3}(1-u)\} - \hat{F}_{Y|2}\{\hat{F}^{-1}_{Y|1}(v)\}, \tag{3.1.1}$$

if $\hat{F}^{-1}_{Y|1}(v) \leq \hat{F}^{-1}_{Y|3}(1-u)$.

The estimated VUS is constructed as

$$\widehat{VUS}(t_0) = \int_0^1 \int_0^1 \hat{Q}(u,v)dudv. \tag{3.1.2}$$

The resulting estimated ROC surface are increasing step functions jumping at event times. We approximate the $\widehat{VUS}(t_0)$ by summation of the volumes of rectangular prisms for their lengths, widths and heights corresponding to the CCP1, CCP2, CCP3 given the varying threshold pairs of (c_1, c_2).

3.2 ASYMPTOTIC PROPERTIES AND INFERENCE OF $\widehat{VUS}(T_0)$

The estimation procedure of $\widehat{VUS}(t_0)$ includes the estimation of the conditional CDF of biomarker Y given the particular disease status and the surface under the ROC curve. Bayes principle allows formulating the conditional distributions $F_{Y|1}(y)$ as

$$\begin{aligned} F_{Y|1}(y) &= P(Y \leq y | D(t_0) = 1) \\ &= \frac{P\left(T_i \leq t_0, \epsilon_i = 1 | Y_i \leq y\right) P\left(Y_i \leq y\right)}{P(T_i \leq t_0, \epsilon_i = 1)}, \end{aligned}$$

where $P(T_i \leq t_0, \epsilon_i = 1 | Y_i \leq y)$ is the conditional CIF of the cause-1 event given the subset $Y_i \leq y$. The same idea is also applied to $F_{Y|2}(y)$and $F_{Y|3}(y)$ and we have

$$F_{Y|2}(y) = \frac{P\left(T_i \leq t_0, \epsilon_i = 2 | Y_i \leq y\right) P\left(Y_i \leq y\right)}{P(T_i \leq t_0, \epsilon_i = 2)}$$

and

$$F_{Y|3}(y) = \frac{P\left(T_i > t_0 | Y_i \leq y\right) P\left(Y_i \leq y\right)}{P(T_i > t_0)}.$$

Counting process and martingale theory (Kalbfleisch and Prentice, 2011) can be utilized in deriving the influence functions for Kaplan-Meier estimators of survival functions and nonparametric estimators of CIF. $\mathbb{I}_{F_{Y/d}(y)}$, the influence functions of $F_{Y|d}(y)$ for $d = 1, 2, 3$, are developed through Taylor's expansion incorporating the elements of the marginal and the conditional CIF (or survival function), as well as the empirical distribution of Y. Hadamard-differentiability and functional delta method are used in constructing the influence function of the surface under ROC curve $\mathbb{I}_{Q(u,v)}$ with respect to its presentation containing quantile function and compound function. The asymptotic normality of $\widehat{VUS}(t_0)$ is stated in the following theorem.

Theorem 3.1: Let $\nu_1 > \inf\{u : F(u) > 0\}$ and $\nu_2 < \sup\{u : S_X(u) > 0\}$ and t_0 in $[\nu_1, \nu_2]$. Subject to independent censoring, the estimated VUS for the ordered competing risks outcomes associated with the prognostic biomarker Y through their correct classification probabilities at t_0 are uniformly consistent and asymptotically normally distributed and we can write

$$\sqrt{n}(\widehat{VUS}(t_0) - VUS(t_0)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{I}_{\widehat{VUS}(t_0)} + o_p(1),$$

where $\mathbb{I}_{\widehat{VUS}(t_0)}$ is the influence function of $\widehat{VUS}(t_0)$.

The work on the proof of Theorem 3.1 is presented below with great details. The estimated variance of $\widehat{VUS}(t_0)$ is

$$\hat{\sigma}^2_{\widehat{VUS}(t_0)} = \frac{1}{n} \sum_{i=1}^{n} \hat{\mathbb{I}}^2_{\widehat{VUS}(t_0)}.$$

Given $\widehat{VUS}(t_0)$ is asymptotic normal, we construct a Wald-type $(1-\alpha)$ confidence interval for $\widehat{VUS}(t_0)$ as

$$\left\{ \widehat{VUS}(t_0) - z_{1-\alpha/2} \frac{\hat{\sigma}_{\widehat{VUS}(t_0)}}{\sqrt{n}}, \widehat{VUS}(t_0) + z_{1-\alpha/2} \frac{\hat{\sigma}_{\widehat{VUS}(t_0)}}{\sqrt{n}} \right\}$$

.

- Proof of Theorem 3.1

  In estimating the variance of $\widehat{VUS}$, we develop its influence function based on the estimating procedures of the conditional CDF of the biomarker Y given the subjects falling into each of the three disease statuses $F_{Y|d}(y)$ and ROC surface $Q(u, v)$. The Bayes theorem formulates $F_{Y|d}(y)$ for $d = 1, 2, 3$ as a ratio of the conditional CIF of the cause-1 event (or the conditional CIF of the cause-2 event, or the conditional survival function) with the CDF of the biomarker Y, over the marginal CIF the cause-1 event (or the marginal CIF of the cause-2 event, or the marginal survival function). The survival function $S(t_0) = P(T > t_0)$ and the cause-specific CIF is defined with association of the survival function and its cause-specific hazard:

$$F_k(t_0) = P(T \leq t_0, \epsilon = k) = \int_0^{t_0} S(s)\, d\Lambda_k,$$

  where $k = 1, 2$, and

$$\Lambda_k(t_0) = \int_0^{t_0} \lambda_k(u)\, du$$

  through

$$\lambda_k(t_0) = \lim_{h \to 0} \frac{P(t_0 \leq T \leq t_0 + h, \epsilon = k \mid T > t_0)}{h}.$$

  Also, let $\Lambda = \Lambda_1 + \Lambda_2$. In real dataset, we observe the triplet (X_i, η_i, Y_i). In counting process formulation, we specify the right-continuous process of the cause-specific events as:

$$N_{ki}(s) = I(X_i \leq s, \eta_i = k),$$

$$N_{k.}(s) = \sum_{i=1}^{n} N_{ki}(s),$$

  as well as the combined events

$$N_i(s) = N_{1i}(s) + N_{2i}(s),$$

$$N_{\cdot}(s) = N_{1.}(s) + N_{2.}(s).$$

  where $I(\cdot)$ is an indicator function. Also, the left-continuous process of the subjects at riskset is

$$R_i(s) = I(X_i \geq s),$$

$$R_{\cdot}(s) = \sum_{i}^{n} R_i(s).$$

The martingale process for either cause-specific event is formulated as

$$M_{ki}(s) = N_{k\cdot}(s) - \int_0^s R_i(u)\, d\Lambda_k.$$

And $M_i(s) = N_{\cdot}(s) - \int_0^s R_i(u)\, d\Lambda$ is another martingale process.

$\hat{S}(t_0)$, the Kaplan-Meier estimator of $S(t_0)$, is

$$\prod_{s \le t_0} \left\{ 1 - \frac{\Delta N_{\cdot}(s)}{R_{\cdot}(s)} \right\}$$

where

$$\Delta N_{\cdot}(s) = N_{\cdot}(s) - N_{\cdot}(s-).$$

The nonparametric estimator of $F_k(t_0)$ is

$$\hat{F}_k(t_0) = \int_0^{t_0} \hat{S}(s)\, d\hat{\Lambda}_k(s),$$

where

$$\hat{\Lambda}_k(s) = \int_0^s \frac{dN_{k\cdot}(s)}{R_{\cdot}(s)}.$$

A body of literature has been devoted to explore their influence functions (Pepe, 1991; Lin, 1997; Zhang and Fine, 2008) for their asymptotic properties. Under regular conditions, the influence functions of the estimators for survival functions and the cumulative incidence functions in the competing risks are

$$\begin{aligned}
\sqrt{n}(\hat{S}(t_0) - S(t_0)) &= -\frac{S(t_0)}{\sqrt{n}} \sum_{i=1}^{n} \int_0^{t_0} \frac{\hat{S}(s-)}{S(s)R_{\cdot}(s)}\, dM_i(s) + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{I}_{S_i}(t_0) + o_p(1);
\end{aligned}$$

$$\begin{aligned}
\sqrt{n}(\hat{F}_k(t_0) - F_k(t_0)) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \int_0^{t_0} \frac{\hat{S}(s-)}{R_{\cdot}(s)}\, dM_{ki}(s) \right. \\
&\quad \left. - \int_0^{t_0} S(s-) \left(\int_0^s \frac{\hat{S}(u-)}{S(u)R_{\cdot}(u)}\, dM_i(u) \right) d\Lambda_k(s) \right\} + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{I}_{F_k i}(t_0) + o_p(1).
\end{aligned}$$

for $k = 1, 2$. The kaplan-Meier estimators and the nonparametric estimators of CIFs are plugged in to substitute $S(t_0)$ and $F_k(t_0)$ in deriving their influence functions.

Let $N_{1i}^y(s)$, $N_{2i}^y(s)$, $N_i^y(s)$, $N_{1\cdot}^y(s)$, $N_{2\cdot}^y(s)$, $N_{\cdot}^y(s)$, $R_i^y(s)$,and $R_{\cdot}^y(s)$ have the same definitions of $N_{1i}(s)$, $N_{2i}(s)$, $N_i(s)$, $N_{1\cdot}(s)$, $N_{2\cdot}(s)$, $N_{\cdot}(s)$, $R_i(s)$, and $R_{\cdot}(s)$ within the subset $Y_i \leq y$ when y follows the distribution of Y. Analogous to the aforementioned estimators in term of the whole sample, we have

$$\hat{S}^y(t_0) = \prod_{s \leq t_0} \left\{ 1 - \frac{\Delta N_{\cdot}^y(s)}{R_{\cdot}^y(s)} \right\},$$

where

$$\Delta N_{\cdot}^y(s) = N_{\cdot}^y(s) - N_{\cdot}^y(s-).$$

Similarly,

$$\hat{F}_k^y(t_0) = \int_0^{t_0} \hat{S}^y(s-) d\hat{\Lambda}_k^y(s),$$

where

$$\hat{\Lambda}_k^y(s) = \int_0^s \frac{dN_{k\cdot}^y(s)}{R_{\cdot}^y(s)},$$

with ^{y} index the estimators or parameters associated with the subset $Y_i \leq y$. The martingale processes for the subset $Y \leq y$ are

$$M_{ki}^y(s) = N_{k\cdot}^y(s) - \int_0^s R_i^y(u) d\Lambda_k^y,$$

Heuristically, $M_i^y(s) = N_{\cdot}^y(s) - \int_0^s R_i^y(u) d\Lambda^y$.

The influence functions of conditional survival function and condition CIF in the subset of $Y_i \leq y$ are modified

$$\begin{aligned} \sqrt{n}(\hat{S}^y(t_0) - S^y(t_0)) &= -\frac{S^y(t_0)}{\sqrt{n}} \sum_{i=1}^n \int_0^{t_0} \frac{\hat{S}^y(s-)}{S^y(s) R_{\cdot}^y(s)} \, dM_i^y(s) + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{I}_{S_i^y}(t_0) + o_p(1); \end{aligned}$$

$$\begin{aligned}\sqrt{n}(\hat{F}_k^y(t_0) - F_k^y(t_0)) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Bigg\{\int_0^{t_0}\frac{\hat{S}^y(s-)}{R_\cdot^y(s)}\,dM_{ki}^y(s) \\ &\quad -\int_0^{t_0} S^y(s-)\left(\int_0^s \frac{\hat{S}^y(u-)}{S^y(u)R_\cdot^y(u)}\,dM_i^y(u)\right)\,d\Lambda_k^y(s)\Bigg\} + o_p(1) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbb{I}_{F_{ki}{}^y}(t_0) + o_p(1),\end{aligned}$$

and $k = 1, 2$.

The influence function of the empirical distribution of CDF $F(y)$ is

$$\begin{aligned}\sqrt{n}(\hat{F}(y) - F(y)) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{I(Y_i \le y) - F(y)\} + o_p(1) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbb{I}_{F_i}(y) + o_p(1).\end{aligned}$$

$\mathbb{I}_{F_i}(y)$ is estimated by plugging in the estimated empirical cumulative density distribution $\hat{F}(y)$ for $F(y)$.

Employing Taylor expansion $f(\mathbb{E}, \mathbb{F}, \mathbb{G}) = \frac{\mathbb{E}\mathbb{F}}{\mathbb{G}}$, we develop

$$\frac{\widehat{\mathbb{E}}\widehat{\mathbb{F}}}{\widehat{\mathbb{G}}} - \frac{\mathbb{E}\mathbb{F}}{\mathbb{G}} = \frac{\mathbb{F}\mathbb{G}(\widehat{\mathbb{E}} - \mathbb{E}) + \mathbb{E}\mathbb{G}(\widehat{\mathbb{F}} - \mathbb{F}) - \mathbb{E}\mathbb{F}(\widehat{\mathbb{G}} - \mathbb{G})}{\mathbb{G}^2} + o_p(1). \tag{3.2.1}$$

For illustration, we show details about derivation of the influence function of the conditional distribution $F_{Y/3}(y)$ as

$$\begin{aligned}&\mathbb{E} = P\left(T_i > t_0 | Y_i \le y\right) = S^y(t_0), \\ &\mathbb{F} = P\left(Y_i \le y\right) = F(y), \\ &\mathbb{G} = P\left(T_i > t_0\right) = S(t_0), \\ &\widehat{\mathbb{E}} - \mathbb{E} = \hat{S}^y(t_0) - S^y(t_0) \approx \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{S_i^y}(t_0), \\ &\widehat{\mathbb{F}} - \mathbb{F} = \hat{F}(y) - F(y) \approx \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{F_i}(y), \\ &\widehat{\mathbb{G}} - \mathbb{G} = \hat{S}(t_0) - S(t_0) \approx \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{S_i}(t_0).\end{aligned}$$

By filling in each elements back to the Taylor expansion equation (3.2.1), we yield

$$\begin{aligned}\sqrt{n}(\hat{F}_{Y|3}(y) - F_{Y|3}(y)) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{F(y)S(t_0)\mathbb{I}_{S_i^y}(t_0) + S^y(t_0)S(t_0)\mathbb{I}_{F_i}(y) - S^y(t_0)F(y)\mathbb{I}_{S_i}(t_0)}{S(t_0)^2} \\ &\quad +o_p(1) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbb{I}_{F_{(Y|3)i}(y)} + o_p(1), \end{aligned} \tag{3.2.2}$$

and we estimate the influence function of $\hat{F}_{Y|3}(y)$ by plugging in $\hat{F}(y), \hat{S}(t_0), \hat{S}^y(t_0)$ for $F(y), S(t_0), S^y(t_0)$.

With the same approach, the influence functions of $\hat{F}_{Y|k}(y)$ for $k = 1, 2$ are written as:

$$\begin{aligned}\sqrt{n}(\hat{F}_{Y|k}(y) - F_{Y|k}(y)) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{F(y)F_k(t_0)\mathbb{I}_{F_{ki}^y}(t_0) + F_k^y(t_0)F_k(t_0)\mathbb{I}_{F_i}(y) - F_k^y(t_0)F(y)\mathbb{I}_{F_{ki}}(t_0)}{F_k(t_0)^2} \\ &\quad +o_p(1) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbb{I}_{F_{(Y|k)i}(y)} + o_p(1), \end{aligned} \tag{3.2.3}$$

Similarly, we estimate the influence function of $\hat{F}_{Y|1}(y)$ and $\hat{F}_{Y|2}(y)$ by plugging in $\hat{F}(y), \hat{F}_1(t_0), \hat{F}_1^y(t_0), \hat{F}_2(t_0), \hat{F}_2^y(t_0)$ for $F(y), F_1(t_0), F_1^y(t_0), F_2(t_0), F_2^y(t_0)$.

The influence function in estimating the surface of ROC curve, $Q(u, v) = F_{Y|2}(F_{Y|3}^{-1}(1 - u)) - F_{Y|2}(F_{Y|1}^{-1}(v))$, if $F_{Y|1}^{-1}(v) \leq F_{Y|3}^{-1}(1 - u)$, is achieved by evaluating the asymptotic properties of the estimators for the quantile function $F_{Y|l}^{-1}(\cdot)$ and for the compound function $F_{Y|2}\left\{F_{Y|l}^{-1}(\cdot)\right\}$ for $l = 1, 3$. The quantile function $F_{Y|l}^{-1}(\tau) = \inf\left(y : F_{Y/l}(y) \geq \tau, \tau \in (0, 1)\right)$ in the competing risks framework, which is the smallest value of y at which the probability of event l exceeding τ in the presence of other competing events. Since F^{-1} is Hadamard-differentiable, using the functional delta method, the influence function of $\hat{F}_{Y|l}^{-1}(\tau)$ is

$$\begin{aligned}\sqrt{n}(\hat{F}_{Y|l}^{-1}(\tau) - F_{Y|l}^{-1}(\tau)) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{-\mathbb{I}_{F_{(Y|l)i}}(y) \circ F_{Y|l}^{-1}(\tau)}{f_{Y|l}(y) \circ F_{Y|l}^{-1}(\tau)} + o_p(1) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbb{I}_{F_{(Y|l)i}^{-1}}(\tau) + o_p(1), \end{aligned} \tag{3.2.4}$$

where $f_{Y|l}(y)$ is the first derivative of $F_{Y|l}(y)$ and $\circ$ is a mathematical operator of function composition. $\mathbb{I}_{F_{(Y|l)i}}$ have been specified in equations (3.2.2) and (3.2.3) respectively. $F^{-1}_{Y|l}(\tau)$ can be estimated with the inverse function of $\hat{F}_{Y|l}(\tau)$, where τ can be replaced by v or $1-u$ corresponding to $l=1$ or 3.

In term of the inference of $\hat{F}_{Y|2}(\hat{F}^{-1}_{Y|l}(\tau))$, we took the idea of asymptotic features of the compound function from quantile association for bivariate survival data (Li et al., 2017). We assume the following regularity conditions:

1. $f_{Y|2}(y)$ is the first derivative of $F_{Y|2}(y)$ and $\left|f_{Y|2}(y)\right|$ is bounded uniformly in (τ_L, τ_U), and τ_L, τ_U are within $(0,1)$ representing the lower and upper bounds of quantile range of interest.
2. $F^{-1}_{Y|l}(\tau)$ is Lipschitz continuous while $\tau \in (\tau_L, \tau_U)$.

We need to express $\hat{F}_{Y|2}(\hat{F}^{-1}_{Y|l}(\tau)) - F_{Y|2}(F^{-1}_{Y|l}(\tau))$ with respect to two presentations to examine its uniform consistency and weak convergence respectively. Firstly in examining uniform consistency of $\hat{F}_{Y|2}(\hat{F}^{-1}_{Y|l}(\tau))$, we write the formula as:

$$\begin{aligned}\sup_{\tau\in(\tau_L,\tau_U)}\left|\hat{F}_{Y|2}(\hat{F}^{-1}_{Y|l}(\tau)) - F_{Y|2}(F^{-1}_{Y|l}(\tau))\right| \leq & \sup_{\tau\in(\tau_L,\tau_U)}\left|\hat{F}_{Y|2}(\hat{F}^{-1}_{Y|l}(\tau)) - F_{Y|2}(\hat{F}^{-1}_{Y|l}(\tau))\right| \\ & + \sup_{\tau\in(\tau_L,\tau_U)}\left|F_{Y|2}(\hat{F}^{-1}_{Y|l}(\tau)) - F_{Y|2}(F^{-1}_{Y|l}(\tau))\right|\end{aligned}$$

We have shown that $\hat{F}_{Y|2}(y)$ is uniformly consistent to $F_{Y|2}(y)$ given equation (3.2.3) and $\hat{F}^{-1}_{Y/l}(\tau)$ is uniformly consistent to $F^{-1}_{Y/l}(\tau)$ given equation (3.2.4). Assembling both evidences, given $|f_{Y|2}(y)|$ as bounded, $\hat{F}_{Y|2}(\hat{F}^{-1}_{Y|l}(\tau)) - F_{Y|2}(F^{-1}_{Y|l}(\tau)) \to 0$ uniformly while $\tau \in (\tau_L, \tau_U)$ as $n \to \infty$.

In developing weak convergence, the equation is written as:

$$\begin{aligned}\sqrt{n}\left(\hat{F}_{Y|2}(\hat{F}^{-1}_{Y|l}(\tau)) - F_{Y|2}(F^{-1}_{Y|l}(\tau))\right) = & \sqrt{n}\left(\hat{F}_{Y|2}(\hat{F}^{-1}_{Y|l}(\tau)) - \hat{F}_{Y|2}(F^{-1}_{Y|l}(\tau))\right) \\ & + \sqrt{n}\left(\hat{F}_{Y|2}(F^{-1}_{Y|l}(\tau)) - F_{Y|2}(F^{-1}_{Y|l}(\tau))\right) \quad (3.2.5)\end{aligned}$$

Under the continuous mapping theorem, $\sqrt{n}\left(\hat{F}_{Y|2}(F^{-1}_{Y|l}(\tau)) - F_{Y|2}(F^{-1}_{Y|l}(\tau))\right)$ converges weakly to a Gaussian process with mean 0 given equation (3.2.3). The influence function of $\hat{F}_{Y|2}(F^{-1}_{Y|l}(\tau))$ is obtained by modifying $\mathbb{I}_{F_{(Y|2)i}(y)}$ with y replaced by $\hat{F}^{-1}_{Y/l}(\tau)$, which we denote as $\mathbb{I}_{F_{(Y|2)i}F^{-1}_{(Y|l)}(\tau)}$.

Furthermore, together with the quantile function $\hat{F}^{-1}_{Y|l}(\tau)$ uniformly converging to $F^{-1}_{Y|l}(\tau)$, given the argument from Peng and Huang (2008), we write

$$\begin{aligned}\sqrt{n}\left(\hat{F}_{Y|2}(\hat{F}^{-1}_{Y|l}(\tau)) - \hat{F}_{Y|2}(F^{-1}_{Y|l}(\tau))\right) &= \sqrt{n}\left(F_{Y|2}(\hat{F}^{-1}_{Y|l}(\tau)) - F_{Y|2}(F^{-1}_{Y|l}(\tau))\right) + o_p(1) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_{Y|2}(F^{-1}_{Y|l}(\tau))\mathbb{I}^{-1}_{F_{(Y|l)i}}(\tau) + o_p(1),\end{aligned}$$

where $f_{Y|2}(F^{-1}_{Y|l}(\tau))$ is the first derivative of the conditional CDF $F_{Y|2}(F^{-1}_{Y|l}(\tau))$ evaluated at $F^{-1}_{Y|l}(\tau)$. $\mathbb{I}^{-1}_{F_{(Y/l)i}}(\tau)$ can be derived from equation (3.2.4). Combining the two compartments of equation (3.2.5), we have

$$\begin{aligned}\sqrt{n}\left(\hat{F}_{Y|2}(\hat{F}^{-1}_{Y/l}(\tau)) - F_{Y|2}(F^{-1}_{Y|l}(\tau))\right) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(f_{Y|2}(F^{-1}_{Y|l}(\tau))\mathbb{I}_{F^{-1}_{(Y|l)i}}(\tau) \\ &\quad + \mathbb{I}_{F_{(Y|2)i}F^{-1}_{(Y|l)}}(\tau)\Big) + o_p(1) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbb{I}_{F_{(Y|2)i}F^{-1}_{(Y|l)i}}(\tau) + o_p(1).\end{aligned}$$

Summarizing the results above, the inference of $\hat{Q}(1-u,v)$ is produced with the simple linear algebra of elements $\mathbb{I}_{F_{(Y|2)i}F^{-1}_{(Y|l)i}}(\tau)$,

$$\begin{aligned}\sqrt{n}\Big(\hat{Q}(1-u,v) - Q(1-u,v)\Big) &= \sqrt{n}\Big(\hat{F}_{Y|2}(\hat{F}^{-1}_{Y|3}(1-u)) - F_{Y|2}(F^{-1}_{Y|3}(1-u))\Big) \\ &\quad -\Big(\hat{F}_{Y|2}(\hat{F}^{-1}_{Y|1}(v)) - F_{Y|2}(F^{-1}_{Y|1}(v))\Big) + o_p(1) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(\mathbb{I}_{F_{(y|2)i}F^{-1}_{(y|3)i}}(1-u) - \mathbb{I}_{F_{(y|2)i}F^{-1}_{(y|1)i}}(v)\Big) + o_p(1) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbb{I}_{Q_i(1-u,v)} + o_p(1).\end{aligned}$$

The influence function of $\widehat{VUS}(t_0)$ is derived by integrating $\mathbb{I}_{Q_i(u,v)}$ over the uniform squares,

$$\begin{aligned}\sqrt{n}\left(\widehat{VUS}(t_0) - VUS(t_0)\right) &= \sqrt{n}\left(\int_0^1\int_0^1 (\hat{Q}(u,v) - Q(u,v))\,du\,dv\right) + o_p(1) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\int_0^1\int_0^1 \mathbb{I}_{Q_i(1-u,v)}\,du\,dv\right) + o_p(1).\end{aligned}$$

Therefore, we have the asymptotic weak convergence of $\widehat{VUS}(t_0)$ over $t_0 \in [\nu_1, \nu_2]$ as defined in Theorem 3.1.

4.0 CONCORDANCE DEFINITION BASED ESTIMATION OF VUS

4.1 IPCW-ADJUSTED ESTIMATION OF VUS ACCORDING TO THE CONCORDANCE DEFINITION

Without independent censoring, the disease status by a fixed time $D(t_0)$ would be observed for each subject in the sample. A U-type VUS estimator can be constructed according to randomly selecting three subjects, each from one type of disease status, to reflect concordance relationship between the biomarker and the outcomes. However, in the presence of independent censoring, there are four possible scenarios for the i-th subject:

$$\begin{aligned} I\{X_i \le t_0, \eta_i = 1\} &= I\{T_i \le t_0, \epsilon_i = 1, C_i \ge T_i\} \\ I\{X_i \le t_0, \eta_i = 2\} &= I\{T_i \le t_0, \epsilon_i = 2, C_i \ge T_i\} \\ I\{X_i > t_0\} &= I\{T_i > t_0, C_i > t_0\} \\ I\{X_i \le t_0, \eta_i = 0\} &= I\{C_i \le t_0, T_i > C_i\}. \end{aligned}$$

The disease status $D(t_0)$ is determinable for the first 3 scenarios, but not for the fourth one. To adjust to missing disease statuses due to independent censoring before t_0, we adopt inverse probability censoring weighting (IPCW). IPCW is used to inversely weight the observed subjects of the cause-1 event and the cause-2 event at the selected t_0 according to their probabilities of being observed at the times of event occurrence, and the weights were estimated based on the censored data. For more information of general theory of IPCW, we can refer to Van der Laan and Robins (2003). Let $G(t) = P(C > t)$ be the survival function of censoring, and $\hat{G}(t)$ be the Kaplan-Meier estimator of $G(t)$. Following that the

Kaplan-Meier estimator is a self-consistent estimator with right-censored data redistribution (Efron (1967); Dinse (1985)), $I\{X_i \leq t_0, \eta_i = k\}/\hat{G}(X_i)$ is an unbiased estimator of $P(T_i \leq t_0, \epsilon_i = k)$ for $k = 1, 2$ and $I\{X_i > t_0\}/\hat{G}(t_0)$ is an unbiased estimator of $P(T_i > t_0)$. Hence, the proposed U-type estimator of VUS adjusted for IPCW is written as:

$$\widetilde{VUS}(t_0) = \frac{\sum_i \sum_{j\neq i} \sum_{k\neq i,j} \frac{I(X_i \leq t_0, \eta_i = 1, X_j \leq t_0, \eta_j = 2, X_k > t_0, Y_i < Y_j < Y_k)}{\hat{G}(X_i-)\hat{G}(X_j-)\hat{G}(t_0)}}{\sum_i \sum_{j\neq i} \sum_{k\neq i,j} \frac{I(X_i \leq t_0, \eta_i = 1, X_j \leq t_0, \eta_j = 2, X_k > t_0)}{\hat{G}(X_i-)\hat{G}(X_j-)\hat{G}(t_0)}}. \tag{4.1.1}$$

As a consequence, the VUS estimator is the ratio of the estimated probability of observing a triplet of subjects each from one of the three disease statuses, with their associated ordered biomarker, over the estimated probability of observing the triplet each from one of the three disease statuses.

4.2 CONSISTENCY AND WEAK CONVERGENCE OF $\widetilde{VUS}(T_0)$

IPCW is used to account for missing data from independent right-censoring, which weights the observed subjects by their inverse probability of being observed, and the weight is estimated from the censored subjects. IPCW intends to recreate a population which would have been seen without censored subjects. In the scope of this study, we focus on biomarker-independent censoring. However, in the presence of biomarker-dependent censoring, without any modification, IPCW-adjusted estimators might cause some bias due to the degree of association between the censoring mechanism and the biomarker. Blanche et al. (2013b) stated that a little modification of inverse probability of censoring weighting can do a similar job to the nearest neighbor estimator which was proposed by Heagerty et al. (2000) for biomarker-dependent censoring. IPCW-adjusted weighting is to weight the observed subjects with the marginal probability of be censored without accounting for the biomarker's association to the censoring, as compared to the modified (conditional) IPCW-adjusted weighting which is to weight the observed subjects at the time of being observed by conditional probability

of being censored given the biomarker. In this research, modified IPCW-adjusted weighting is not in our research scope, but it is straightforward to incorporate modified-IPCW distribution to accommodate a biomarker-dependent censoring given our proposed method.

For now, let $\hat{G}(t)$ be the Kaplan-Meier estimator of survival function of censoring $G(t) = P(C > t)$. As $n \to \infty$, $\frac{1}{n}\sum_{i=1}^{n} I(X_i \leq t_0, \eta = 1)/\hat{G}(X_i)$ converge to

$$\begin{aligned} E\left\{\frac{I(X_i \leq t_0, \eta_i = 1)}{G(X_i)}\right\} &= E\left\{E\left[\frac{I(T_i \leq t_0, \epsilon_i = 1)I(T_i \leq C_i)}{G(X_i)} \mid X_i, \eta_i\right]\right\} \\ &= E\left\{I(T_i \leq t_0, \epsilon_i = 1)\left[\frac{E\{I(T_i \leq C_i)|X_i, \eta_i\}}{G(X_i)}\right]\right\} \\ &= P(T_i \leq t_0, \epsilon_i = 1). \end{aligned}$$

By analogy,

$$\frac{1}{n}\sum_{i=1}^{n} I(X_j \leq t_0, \eta_j = 2)/\hat{G}(X_j)$$

and

$$\frac{1}{n}\sum_{i=1}^{n} I(X_k > t_0)/\hat{G}(t_0)$$

converge to $P(T_j \leq t_0, \epsilon_j = 2)$ and $P(T_k > t_0)$ respectively.

Base on the proof that the IPCW-adjusted probability of observing those with the cause-1 event (or the cause-2 event or none of the events) by a fixed time t_0 is a consistent to the probability of subjects experiencing the cause-1 event (or the cause-2 event or none of the events) by t_0, heuristically, with Slutsky's theorem, we have

$$\begin{aligned} E(\widetilde{VUS}(t_0)) &= E\left\{E\left[\frac{\frac{I(X_i \leq t_0, \eta_i = 1)}{G(X_i-)}\frac{I(X_j \leq t_0, \eta_j = 2)}{G(X_j-)}\frac{I(X_k > t_0)}{G(t_0)}I(Y_i < Y_j < Y_k)}{\frac{I(X_i \leq t_0, \eta_i = 1)}{G(X_i-)}\frac{I(X_j \leq t_0, \eta_j = 2)}{G(X_j-)}\frac{I(X_k > t_0)}{G(t_0)}} \mid \mathbb{Q}\right]\right\} \\ &= E\left(\frac{I(T_i \leq t_0, \epsilon_i = 1, T_j \leq t_0, \epsilon_j = 2, T_k > t_0, Y_i < Y_j < Y_k)}{I(T_i \leq t_0, \epsilon_i = 1, T_j \leq t_0, \epsilon_j = 2, T_k > t_0)}\right) \\ &= P(Y_1 < Y_2 < Y_3 | T_i \leq t_0, \epsilon_i = 1, T_j \leq t_0, \epsilon_j = 2, T_k > t_0). \end{aligned}$$

Given the observed dataset $\mathbb{Q} = (X_i, X_j, X_k, \eta_i, \eta_j, Y_i, Y_j, Y_k)$, the IPCW-adjusted $\widetilde{VUS}(t_0)$ leads to a consistent estimator of VUS.

Weak convergence of $\widetilde{VUS}(t_0)$ using counting process techniques and the theory of U-statistics results in an explicit formula for the variance estimation, where we adapted the

proof from Hung and Chiang (2010) for survival data without competing risks events. For a fixed t_0, the asymptotic result of the Kaplan-Meier estimator has its martingale expression for the censoring survival function $G(t_0)$ as $n \to \infty$:

$$\sup \left| \sqrt{n}(\hat{G}(t_0) - G(t_0)) + \frac{G(t_0)}{\sqrt{n}} \sum_{i=1}^{n} \int_0^{t_0} \frac{dM_{C_i}(u)}{S_X(u)} \right| = o_p(1).$$

And it can be rewritten as the ratio of

$$\frac{\hat{G}(t_0)}{G(t_0)} - 1 = -\frac{1}{n} \sum_{i=1}^{n} \int_0^{t_0} \frac{dM_{C_i}(u)}{S(u)} + o_p(1). \tag{4.2.1}$$

Note that

$$M_{C_i}(t_0) = I(\eta_i = 0, X_i \leq t_0) - \int_0^{t_0} I(X_i \geq t_0) d\Lambda_C(u)$$

with $\Lambda_C(\cdot)$ being a cumulative hazard function of the censoring time C. Defining

$$\Lambda_C(t_0) = \int_0^{t_0} \lambda_C(u)\, du,$$

where

$$\lambda_C(t_0) = \lim_{h \to 0} \frac{P(t_0 \leq C \leq t_0 + h \mid C > t_0)}{h},$$

and $\Lambda_C(\cdot)$ is estimated by plugging in Nelson-Aalen estimators. Similarly, $\hat{S}(t_0)$ is an empirical Kaplan-Meier estimator for the survival function $S(t_0) = P(X > t_0)$.

Theorem 4.1: Given that the censoring variable C is independent of T and t_0 is defined in Theorem 3.1, according to the definition (2.3.1) of VUS, we have

$$\sqrt{n}(\widetilde{VUS}(t_0) - VUS(t_0)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{I}_{\widetilde{VUS}(t_0)} + o_p(1),$$

where $\mathbb{I}_{\widetilde{VUS}(t_0)}$ is the influence function of $\widetilde{VUS}(t_0)$ and $E\left[\mathbb{I}_{\widetilde{VUS}}(X_i, \eta_i, Y_i, t_0)\right] = 0$.

- Proof of Theorem 4.1

  We define

$$\tilde{\mathbb{A}}_{ijk} = \frac{I(X_i \le t_0, \eta_i = 1, X_j \le t_0, \eta_j = 2, X_k > t_0, Y_i > Y_j > Y_k)}{G(X_i)G(X_j)G(t_0)}$$

  where the estimated $\widehat{\mathbb{A}}_{ijk}$ is derived by plugging in Kaplan-Meier estimators of $\hat{G}(\cdot)$ to $G(\cdot)$. Let $\mathbb{A} = E(\tilde{\mathbb{A}}_{ijk})$.

  Let $\mathbb{C}_{ijk} = I(X_i \le t_0, \eta_i = 1, X_j \le t_0, \eta_j = 2, X_k > t_0, Y_i > Y_j > Y_k)$. Using the Taylor expansion for $f(x_i, x_j, x_k) \to \frac{\mathbb{C}}{x_i x_j x_k}$, we have

$$\begin{aligned}\frac{\mathbb{C}}{\hat{G}(X_i)\hat{G}(X_j)\hat{G}(t_0)} &= \frac{\mathbb{C}}{G(X_i)G(X_j)G(t_0)}\Big(1 - (\frac{\hat{G}(X_i)}{G(X_i)} - 1) \\ &\quad -(\frac{\hat{G}(X_j)}{G(X_j)} - 1) - (\frac{\hat{G}(t_0)}{G(t_0)} - 1)\Big) + o_p(1)\end{aligned}$$

  By invoking the martingale expression from equation (4.2.1),as well as estimation theory of U statistics, we have

$$\begin{aligned}\sqrt{n}(\hat{\mathbb{A}} - \mathbb{A}) = \Bigg[\frac{\sqrt{n}}{\binom{n}{6}} \sum_{i\neq j\neq k\neq p\neq q\neq r} \widetilde{\mathbb{A}}_{ijk}\Bigg\{1 + \int_0^{X_i} \frac{dM_{C_p}(u)}{S(u)} \\ + \int_0^{X_j} \frac{dM_{C_q}(u)}{S(u)} + \int_0^{t_0} \frac{dM_{C_r}(u)}{S(u)}\Bigg\} - \mathbb{A}\Bigg] + o_p(1).\end{aligned}$$

  $\binom{n}{m}$ is a combination expression of m-element subset from the set of $1, \cdots, n$.

  Similarly, we define

$$\tilde{\mathbb{B}}_{ijk} = \frac{I(X_i \le t_0, \eta_i = 1, X_j \le t_0, \eta_j = 2, X_k > t_0)}{G(X_i)G(X_j)G(t_0)}$$

  and develop

$$\begin{aligned}\sqrt{n}(\hat{\mathbb{B}} - \mathbb{B}) &= \Bigg[\frac{\sqrt{n}}{\binom{n}{6}} \sum_{i\neq j\neq k\neq p\neq q\neq r} \widetilde{\mathbb{B}}_{ijk}\Bigg(1 + \int_0^{X_i} \frac{dM_{C_p}(u)}{S(u)} \\ &\quad + \int_0^{X_j} \frac{dM_{C_q}(u)}{S(u)} + \int_0^{t} \frac{dM_{C_r}(u)}{S(u)}\Bigg) - \mathbb{B}\Bigg] + o_p(1).\end{aligned}$$

Applying another Taylor's expansion for $f(x,y) \to \frac{x}{y}$, we have

$$\frac{\hat{\mathbb{A}}}{\hat{\mathbb{B}}} - \frac{\mathbb{A}}{\mathbb{B}} = \frac{(\hat{\mathbb{A}} - \mathbb{A}) - \frac{\mathbb{A}}{\mathbb{B}}(\hat{\mathbb{B}} - \mathbb{B})}{\mathbb{B}} + o_p(1).$$

Assembling the aforementioned results, we derive

$$\sup_{t_0} \left| \sqrt{n}(\widetilde{VUS} - VUS) - \frac{\sqrt{n}}{\binom{n}{6}} \sum_{i \neq j \neq k \neq p \neq q \neq r} \Psi_{ijkpqr}(t_0) \right| = o_p(1), \tag{4.2.2}$$

where

$$\begin{aligned}
\Psi_{ijkpqr}(t_0) \quad = \quad & \frac{1}{\mathbb{B}} \Bigg\{ \widetilde{\mathbb{A}}_{ijk} \big(1 + \int_0^{X_i} \frac{dM_{C_p}(u)}{S(u)} \\
& + \int_0^{X_j} \frac{dM_{C_q}(u)}{S(u)} + \int_0^{t_0} \frac{dM_{C_r}(u)}{S(u)}\big) - \mathbb{A} \\
& - \frac{\mathbb{A}}{\mathbb{B}} \big[\widetilde{\mathbb{B}}_{ijk} \big(1 + \int_0^{X_i} \frac{dM_{C_p}(u)}{S(u)} \\
& + \int_0^{X_j} \frac{dM_{C_q}(u)}{S(u)} + \int_0^{t_0} \frac{dM_{C_r}(u)}{S(u)}\big) - \mathbb{B} \big] \Bigg\}.
\end{aligned}$$

U statistics theory (Lee, 1990) is a powerful tool in studying large sample properties. Belonging to a non-parametric family, U statistics usually are uniformly minimum variance unbiased estimators (UMVUE) for the parameter θ. For a distribution family $\mathfrak{F}$ with the parameter θ and a sample $(X_1, \ldots, X_n)$, the definition of U statistics is

$$U_n = \binom{n}{m}^{-1} \sum h(X_{i1}, \ldots, X_{im}), \tag{4.2.3}$$

where $\sum$ denotes the summation over the $\binom{n}{m}$ combinations of m distinct elements $(X_{i1}, \ldots, X_{im})$ from the sample $(1, \ldots, n)$. A Borel function $h(\cdot)$ is called a kernel function which is symmetric. In Hoeffding's theorem, the variance of the U-statistics given by definition 4.2.3 with $E[h(X_1, \ldots, X_m)]^2 < \infty$ is written,

$$Var(U_n) = \binom{n}{m}^{-1} \sum_{k=1}^{m} \binom{m}{k} \binom{n-m}{m-k} \zeta_k,$$

where

$$\zeta_k = Var[h_k(X_1, ..., X_k)].$$

An important Corollary from Hoeffding's theorem is developed as

$$\frac{m^2}{n}\zeta_1 \leq Var(U_m) \leq \frac{m}{n}\zeta_m.$$

We apply the Hájek projection theory and Hoeffding's decomposition of U statistics (Van Der Vaart, 1998) and derive:

$$\frac{\sqrt{n}}{\binom{n}{6}} \sum_{i\neq j\neq k\neq p\neq q\neq r} \Psi_{ijkpqr}(t_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{I}_{\widetilde{VUS}}(X_i, \eta_i, Y_i, t_0) + o_p(1),$$

where accounting for permutation symmetry of U statistics, we can write $\mathbb{I}_{\widetilde{VUS}(t_0)}$ as projection of the kernel function of degree 6, defining as

$$\begin{aligned} h_{ijkpqr}(t_0) &= \frac{1}{6!} \sum_{i\neq j\neq k\neq p\neq q\neq r} (\Psi_{ijkpqr}(t_0) + \Psi_{jikpqr}(t_0) + \Psi_{jkipqr}(t_0) + \Psi_{jkpiqr}(t_0) \\ &\quad + \Psi_{jkpqir}(t_0) + \Psi_{jkpqri}(t_0)), \end{aligned}$$

corresponding to its conditional expectation on the random vector (X_i, η_i, Y_i),

$$\begin{aligned} \mathbb{I}_{\widetilde{VUS}}(X_i, \eta_i, Y_i, t_0) &= E(\Psi_{ijkpqr}(t_0) + \Psi_{jikpqr}(t_0) + \Psi_{jkipqr}(t_0) + \Psi_{jkpiqr}(t_0) + \Psi_{jkpqir}(t_0) \\ &\quad + \Psi_{jkpqri}(t_0) \mid (X_i, \eta_i, Y_i)). \end{aligned}$$

As a centered projection,$E(\mathbb{I}_{\widetilde{VUS}}(X_i, \eta_i, Y_i, t_0)) = 0$.

Furthermore, the variance of $\widetilde{VUS}(t_0)$ can be consistently estimated by

$$\hat{\sigma}^2_{\widetilde{VUS}(t_0)} = \frac{1}{n} \sum_{i=1}^{n} \hat{\mathbb{I}}_{\widetilde{VUS}}(X_i, \eta_i, Y_i, t_0)^2.$$

Weak convergence for $\widetilde{VUS}(t_0)$ over $t_0 \in [\nu_1, \nu_2]$ follows naturally.

One exciting contribution associated with our development of this methodology is that we have implemented R code in estimating the variance of $\widetilde{VUS}(t_0)$. The matrix operations are well formulated in R program based on their mathematical expressions including martingale process for censoring and projection theory of U-statistics with a 6-degree kernel function. The R codes are available once this research-related paper is accepted, with which those interested users can directly obtain the estimator as well as its estimated standard error without time-consuming bootstrapping.

The Wald-type $(1-\alpha)$ confidence interval of $\widetilde{VUS}(t_0)$ is

$$\left\{ \widetilde{VUS}(t_0) - z_{1-\alpha/2} \frac{\hat{\sigma}_{\widetilde{VUS}(t_0)}}{\sqrt{n}}, \widetilde{VUS}(t_0) + z_{1-\alpha/2} \frac{\hat{\sigma}_{\widetilde{VUS}(t_0)}}{\sqrt{n}} \right\}.$$

5.0 TIED SCORES IN BIOMARKER AND HYPERVOLUME UNDER THE ROC MANIFOLD FOR HIGH-DIMENSIONAL OUTCOMES

5.1 ADAPTION OF TIED SCORES IN THE BIOMARKER

We often encounter ties in a biomarker, for example it is rounded to nearest integers. To account for ties in the biomarker, we modify the formula of VUS in definition (2.3.1) by replacing $I(Y_i < Y_j < Y_k)$ with a $I(Y_i < Y_j < Y_k) + \frac{1}{2}I(Y_i < Y_j = Y_k) + \frac{1}{2}I(Y_i = Y_j < Y_k) + \frac{1}{6}I(Y_i = Y_j = Y_k)$, similarly to the ideas in Wang and Cheng (2014). With the linear interpolation in the biomarker, the asymptotic properties still hold such as uniform consistency and weak convergence. The weights correspond to how the tied scores can contribute in predicting outcome events with a half probability when there are two tied scores, or contribute with one sixth probability when Y_i, Y_j, Y_k appear to be tied simultaneously. Note that the ties occurring in the same disease status have no impact on their concordance association.

With respect to $\widehat{VUS}$ in handling tied scores in a biomarker, we can obtain insight through a geometry exercise in ROC curve (Horton, 2016). We observe that the ROC curve is an exact step function of an ordered biomarker when there are no ties. Either sensitivity or specificity changes when the ordered biomaker moves from one value to the next, as a unique value of the biomarker is associated with either a case or a control. AUC is to summarize accuracy by adding up rectangles corresponding to the area under ROC curve while Y moves through all possible values in order. However, in the presence of tied scores in the biomarker, when a single tied score is associated with both cases and controls, it leads to change of sensitivity and specificity simultaneously as the biomarker moves to the next value, causing a sloped line segment in the ROC curve. Thus, the AUC can be under-

or over-estimated depending on how tied scores affect concordance between the underneath true continuous scores and their outcomes. Excitingly, the practice in AUC can be carried over to our proposed VUS. Again the volume under the ROC surface can be under- or over-estimated depending on how tied scores affect concordance (discordance) relationships. In real clinical studies, it is reasonable to assume that tied records appear in a random pattern without any systematic trend. So $\widehat{VUS}$ is robust against tied scores in a biomarker. Later we show through simulations by rounding Y into tied scores that the proposed $\widetilde{VUS}$ and $\widehat{VUS}$ can handle a tied biomarker well.

To better illustrate the phenomenon of a tied biomarker affecting AUC, we provide an example of a two-category outcome (1 = diseased group and 0 = healthy group). We have a sample of 20 subjects with their biomarkers Y taking distint values $(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20)$ in order, which correspond to 20 binary outcomes $(1, 1, 1, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0)$. By assuming the lower values of Y indicating the greater likelihood of being diseased, it leads to an estimated concordance association AUC 0.83. Then, we examine how the tied scores in the biomarker affect AUC. First, we replace the scores of Y at positions 9 and 10 with their average 9.5. Note that the pair of untied scores in positions 9 and 10 associated with a case and a control, which manifests a concordance between the pair of markers and their outcomes. As a consequence, the tied scores reduce the value of AUC to 0.825 since the ties diminish the concordance association, which attenuate sensitivity and specificity correspondingly. Next, we replace the scores of Y at positions 8 and 9 with their average 8.5, where the two untied scores are associated with a control and a case respectively, showing a discordance relationship. Therefore, the resulting AUC from tied markers increases to 0.835 due to the tied scores covering the discordance relationship. The following table 5.1 shows the associations of disease status and the original untied scores vs. replaced tied scores under those two scenarios affecting the estimation of AUCs in details.

Table 5.1: Untied scores vs. tied scores in biomarker affecting estimated AUC (Scenario 1 = Replace with ties in Y to mask original concordance relationship; Scenario 2 = Replace with ties in Y to mask original discordance relationship)

Disease status based on Survial outcome (Class 1: $T \leq t_0$; Class 0: $T > t_0$)	Untied scores of biomarker Y	Ties scores in Y (Scenario 1)	Ties scores in Y (Scenario 2)
1	1	1	1
1	2	2	2
1	3	3	3
1	4	4	4
0	5	5	5
1	6	6	6
1	7	7	7
0	8	8	8.5
1	9	9.5	8.5
0	10	9.5	10
1	11	11	11
0	12	12	12
1	13	13	13
0	14	14	14
0	15	15	15
1	16	16	16
0	17	17	17
0	18	18	18
0	19	19	19
0	20	20	20
Estimated AUC	0.83	0.825	0.835

Figure 5.1 displays the ROC curves from the untied scores and tied scores and how the estimated AUC changes by affecting sensitivity and specificity. More specifically, the three figures plot sensitivity (TPR) vs 1-specificity (FPR) in terms of untied scores and tied scores that mask the original concordance (or discordance) relationship to their outcome events. The top left one is the ROC curve from untied scores, the top right is the ROC curve from tied scores to mask the original strength of discordance, and the bottom left one is the ROC curve from tied scores to mask the original strength of concordance. Note that the spaces highlighted in red represent the affected sensitivity and specificity where a tied biomarker is associated with a case and a control simultaneously.

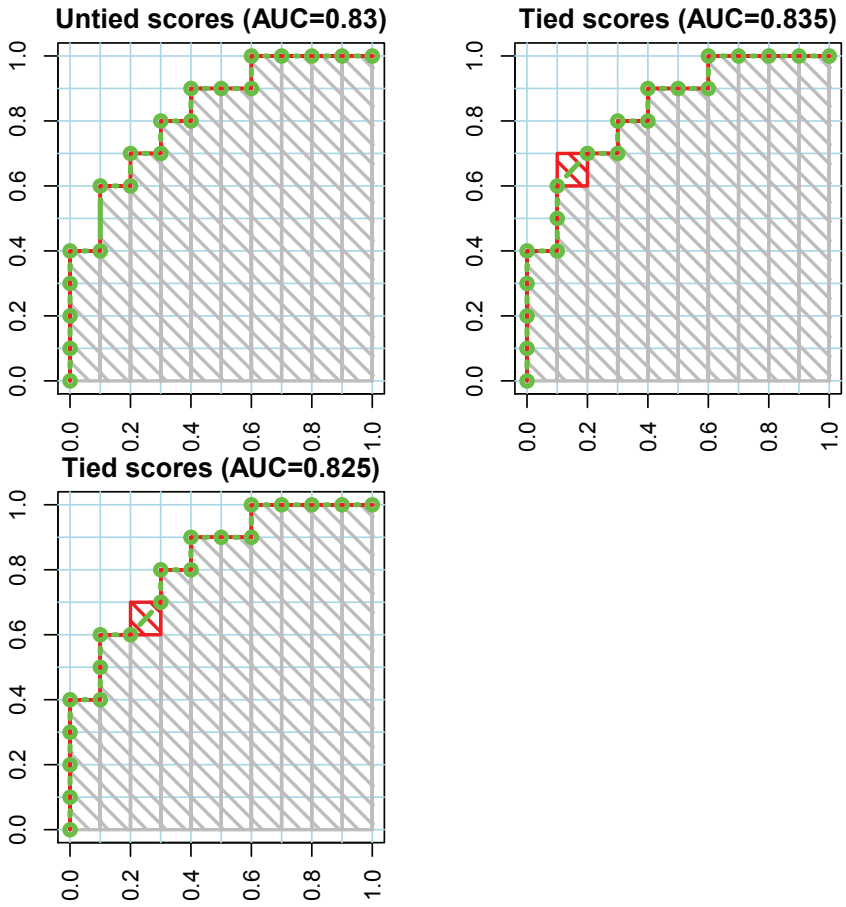

Figure 5.1: ROC curves from untied scores, tied scores which mask the original strength of discordance, and tied scores which mask the original strength of concordance

Furthermore, we take another example to illustrate the three-dimensional ROC surface. By analogy, we generate continuous, untied scores of Y from 1 to 30 along with their associated outcome events. The tied scores were generated by replacing the scores at positions 7, 8, and 9 with the same score 8.5, where the original scores of Y are in a concordance with the outcome statuses. By contrast, the tied scores by replacing the scores in positions 9, 10, and 11 with 10.5 are generated accounting to the original discordance relationship. Table 5.2 shows the estimated $\widehat{VUS}$s from the tied scores are affected by the original concordance and discordance relationships of the underneath untied scores to their outcomes.

Table 5.2: Untied scores vs. tied scores in biomarker affecting estimated VUS (Scenario 1 = Replace with ties in Y to mask original concordance relationship; Scenario 2 = Replace with ties in Y to mask original discordance relationship)

Competing outcome event disease status (Class 1: $T \leq t_0, \epsilon = 1$; Class 2: $T \leq t_0, \epsilon = 2$; Class 3: $T > t_0$)	Untied scores of biomarker Y	Ties scores in Y (Scenario 1)	Ties scores in Y (Scenario 2)
1	1	1	1
1	2	2	2
1	3	3	3
1	4	4	4
1	5	5	5
1	6	6	6
1	7	8.5	7
2	8	8.5	8
3	9	8.5	10.5
2	10	10	10.5
1	11	11	10.5
1	12	12	12
2	13	13	13
2	14	14	14
2	15	15	15
2	16	16	16
1	17	17	17
3	18	18	18
2	19	19	19
2	20	20	20
2	21	21	21
3	22	22	22
2	23	23	23
3	24	24	24
3	25	25	25
3	26	26	26
3	27	27	27
3	28	28	28
3	29	29	29
3	30	30	30
Estimated VUS	0.698	0.693	0.707

5.2 GENERALIZATION TO HUM IN MULTI-WAY ORDINAL OUTCOME EVENTS

It is straightforward to extend the concept of ROC surface and volume under ROC surface for three-category outcome events to multi-way ROC analysis (Scurfield,1998) and hypervolume under the ROC manifold (HUM).

In exploring a medical progression with M-category ordinal outcome events, where $M > 3$, we have M-1 sequentially progressive disease statuses, as compared to healthy controls which is denoted as disease status M. Analogously, HUM corresponds to a probability that the biomarker values and the ordinal outcome events are in concordance. Let $\mathbb{Y}_1, \mathbb{Y}_2, \ldots, \mathbb{Y}_M$ denote the biomarkers from the subjects each from one of the M disease categories corresponding to $\mathbb{T}_1 \leq t_0, \epsilon_1 = 1, \mathbb{T}_2 \leq t_0, \epsilon_2 = 2, \ldots, \mathbb{T}_M > t_0$ respectively. We have the relationship

$$HUM = P(\mathbb{Y}_1 < \mathbb{Y}_2 < \cdots < \mathbb{Y}_M \big| \mathbb{T}_1 \leq t_0, \epsilon_1 = 1, \mathbb{T}_2 \leq t_0, \epsilon_2 = 2, \cdots, \mathbb{T}_M > t_0).$$

Also, subject to independent right-censoring $\mathbb{C}$, we observe $\mathbb{X} = min(\mathbb{T}, \mathbb{C})$ and $\eta = I(\mathbb{T} \leq \mathbb{C})\epsilon$.To account for the data with independent censoring, the IPCW-adjusted $\widetilde{HUM}(t_0)$ by plugging in the Kaplan-Meier estimators $\hat{G}(t)$ for $G(t)$ is written as:

$$\widetilde{HUM}(t_0) = \frac{\sum_1 \sum_2 \cdots \sum_M \frac{I(\mathbb{X}_1 \leq t_0, \eta_1 = 1, \mathbb{X}_2 \leq t_0, \eta_2 = 2, \ldots, \mathbb{X}_M > t_0, \mathbb{Y}_1 < \mathbb{Y}_2 < \cdots < \mathbb{Y}_M)}{\hat{G}(\mathbb{X}_1)\hat{G}(\mathbb{X}_2)\ldots\hat{G}(t_0)}}{\sum_1 \sum_2 \cdots \sum_M \frac{I(\mathbb{X}_1 \leq t_0, \eta_1 = 1, \mathbb{X}_2 \leq t_0, \eta_2 = 2, \ldots, \mathbb{X}_M > t_0)}{\hat{G}(\mathbb{X}_1)\hat{G}(\mathbb{X}_2)\cdots\hat{G}(t_0)}}.$$

Heuristically, correct classification probability can be used as a building component in constructing HUM. Let $\mathbb{F}_{\mathbb{Y}|m}(y)$ denote the conditional CDF of $\mathbb{Y}$ given the subjects in the mth class, where $m \in (1, \ldots, M)$. Similarly, in order to assess the degree of accuracy on the biomarker's ability in predicting the M-catogory events outcomes and in accordance with $M-1$ threshold points $(c_1, c_2, \cdots, c_{M-1})$ with $c_1 < c_2 < \cdots < c_{M-1}$ from the distribution of the biomarker Y, we assign a subject to Class 1 if their biomarker $Y \leq c_1$, to Class 2 if $c_1 < Y \leq c_2$,..., to Class M-1 if $c_{M-2} < Y \leq c_{M-1}$,and to Class M for the remaining.

The correct classification probabilities (CCPs) for the subjects experiencing each of the M-category events by t_0 arw defined as:

$$\begin{aligned} CCP_1 &= P\left(\mathbb{Y}_i \leq c_1 | T_i \leq t_0, \epsilon_i = 1\right) = F_{\mathbb{Y}|1}(c_1), \\ CCP_2 &= P\left(c_1 < \mathbb{Y}_i \leq c_2 | T_i \leq t_0, \epsilon_i = 2\right) = F_{\mathbb{Y}|2}(c_2) - F_{\mathbb{Y}|2}(c_1), \\ &\vdots \\ CCP_M &= P\left(\mathbb{Y}_i > c_{M-1} | T_i > t_0\right) = 1 - F_{\mathbb{Y}|M}(c_{M-1}), \end{aligned}$$

As a consequence, the corresponding $\mathbb{F}_{\mathbb{Y}|m}(y)$s are used for mathematical expressions for each of the CCPs.

Denoting $u = (u_1, \ldots, u_{M-1})$ and following the definitions of Li and Zhou (2009), we have

$$\begin{aligned} c_1 &= F_1^{-1}(u_1), \\ c_2 &= F_2^{-1}(F_2(c_1) + u_2), \\ &\vdots \\ c_{M-1} &= F_{M-1}^{-1}(F_{M-1}(c_{M-2}) + u_{M-1}). \end{aligned}$$

In the M-dimension space, a domain of ROC manifold can be expressed as:

$$\mathbb{Q}(u) = 1 - \mathbb{F}_{\mathbb{Y}|M}(c_{M-1}). \tag{5.2.1}$$

Akin to a three-category competing risks setting, $\hat{\mathbb{F}}_{\mathbb{Y}|m}$, $m = 1, \cdots, M$, are obtained from the nonparametric estimators for bivariate CIF, where continuous Y is treated as a special case of survival data without censoring, and for univariate CIF respectively. Furthermore, plugging in those estimated $\hat{\mathbb{F}}_{\mathbb{Y}|m}(y)$ in into definition 5.2.1, we derive $\hat{\mathbb{Q}}$.

Integrating the possible $CCP_1, CCP2, \ldots, CCP_{M-1}$ in terms of the domain of ROC manifold, we have

$$\widehat{HUM}(t_0) = \int_0^1 \int_0^1 \cdots \int_0^1 \hat{\mathbb{Q}}(u) du_1 du_2 \cdots du_{M-1}.$$

For a M-class outcome, HUM is equal to $1/M!$ when we conduct a random test without any predictive power where $M! = M \times (M-1) \times \cdots \times 1$.

The asymptotic properties of $\widehat{HUM}(t_0)$ and $\widehat{\widetilde{HUM}}(t_0)$ can be established the same as those in $\widehat{VUS}(t_0)$ and $\widehat{\widetilde{VUS}}(t_0)$. Heuristically, the previous discussion on handling tied scores in the biomarker for three categorical outcome events is straightforward to be carried over to the multi-way outcomes events.

6.0 SIMULATION STUDIES

6.1 SIMULATION SETTING

In this section, we conducted simulations to assess the performance of our proposed estimators of VUS from the two different estimating methods. We generated a continuous variable for a biomarker through a uniform distribution as $Y \sim uniform(0, U]$, and specified two points (c_1, c_2) where $0 < c_1 < c_2 < U$, breaking up Y into three segments $(0, c_1], (c_1, c_2], (c_2, U]$ with equal sample sizes in each subset. We generated 3 sets of pairs where $T_d \sim e^{-(\beta_d+\gamma_d)}$ and $\phi_d \sim bin(\frac{\beta_d}{\beta_d+\gamma_d})$, $d = 1, 2, 3$, representing the time to events and the types of events, corresponding to the 3 subsets of Y, where $\epsilon = I(\phi = 1) + 2I(\phi = 0)$. The parameters β_d and γ_d determine the underlying association between the cause-1 event, the cause-2 event and the biomarker Y through their piecewise constant hazard functions

$$\lambda_1(t) = e^{\beta_1 I(Y \le c_1) + \beta_2 I(c_1 < Y \le c_2) + \beta_3 I(Y > c_2)}$$

$$\lambda_2(t) = e^{\gamma_1 I(Y \le c_1) + \gamma_2 I(c_1 < Y \le c_2) + \gamma_3 I(Y > c_2)}.$$

where λ_1 and λ_2 are the hazard rates of the cause-1 event and the cause-2 event, respectively. In the following simulations, we always assigned $U = 20$ and let $c_1 = 5, c_2 = 10$ to break the biomarker Y into the 3 segments $(0, 5], (5, 10]$ and $(10, 20]$. Also, we generated an independent censoring C from a uniform distribution.

We evaluated the performance of the estimators under various scenarios, including three setting of predictive power of the biomarker to the competing events using three scenarios. The first scenario represented a strong predictive association between the two by defining

$\beta_1 = 50$ and $\gamma_1 = 3$ when $Y \in (0, 5]$; $\beta_2 = 3$ and $\gamma_2 = 50$ when $Y \in (5, 10]$; and $\beta_3 = 1$ and $\gamma_3 = 1$ when $Y \in (10, 20]$. Hence, the ratio of occurrences of the cause-1 event to the cause-2 event was 16.7 when the values of biomarker $Y \in (0, 5]$. Then the ratio was reversed in favor of occurrences of the cause-2 event when Y increased to (5, 10], followed by both events declining to the same low rate when Y was greater than 10. The second scenario simulated a moderate predictive association by defining $\beta_1 = 8.0$ and $\gamma_1 = 3.5$; $\beta_2 = 3.5$ and $\gamma_2 = 3.0$; and $\beta_3 = 1.0$ and $\gamma_2 = 2.0$. The third scenario was for the null hypothesis of a completely random test. We specified $\beta_d = 3$ and $\gamma_d = 3$, $d = 1, 2, 3$, with the same hazard rates between the cause-1 event and the cause-2 event across the full spectrum of Y. For each strength of predictive power, we considered three different t_0s, where t_0s were selected with $25 - 30\%$, $65 - 70\%$, and $75 - 85\%$ subjects experienced the events. Also, the censoring rates were 15% and 30% generated from uniform distributions, respectively. We chose sample size = 150, 300 with 1000 simulated datasets for each combination of scenarios.

We computed both $\widetilde{VUS}$ and $\widehat{VUS}$ for each simulated dataset. The estimated IPCW adjusted standard errors $\hat{\sigma}_{\widetilde{VUS}}$ were calculated based on the developed formula. For $\widehat{VUS}(t_0)$, its variance can be estimated from the influence function in Theorem 3.1. However, the evaluation of the influence function is rather complicated. Bootstrapping with 250 replications was used instead to the estimate standard error $\hat{\sigma}_{\widetilde{VUS}}$. We compared the resulting estimators with the true VUS, which was derived based on a large sample without independent censoring (n=300,000). We reported the results in order as follows: the bias $B_{\widetilde{VUS}}$, $ESE_{\widetilde{VUS}}$ (empirical standard error), $ASE_{\widetilde{VUS}}$ (model-based standard error), and $CP_{\widetilde{VUS}}$ (coverage probability) for $\widetilde{VUS}(t_0)$, as well as the bias $B_{\widehat{VUS}}$, $ESE_{\widehat{VUS}}$ (empirical standard error), $BSE_{\widehat{VUS}}$ (bootstrap standard error), and $CP_{\widehat{VUS}}$ (coverage probability) for $\widehat{VUS}(t_0)$.

To assess tied scores in a biomarker, we tested under the scenarios above with a sample size of 300. We generated tied scores of the biomarker on an discrete scale by rounding $Y \in (0, 5]$ or $(10, 20]$ to the nearest 0.1, and rounding $Y \in (5, 10]$ to the nearest 0.2. This design emulated real clinical studies for some ranges of values with higher frequencies than others by their biological natures.

6.2 SIMULATION RESULTS

Table 6.1 outlined the simulation results combining all scenarios for sample size = 300 with 1000 replications. As expected, the mean values from both $\widetilde{VUS}$ and $\widehat{VUS}$ were barely different from the true values, and their coverage rates were close to the nominal level 0.95. The model-based standard error, the bootstrap standard error, and the empirical stanard errors from the two estimators all agree with each other. When there is no association between the biomarker and the three disease statuses under the null hypothesis, the means of the two estimators barely deviate from the true value 0.167 and the coverage rates are very close to 95%. Since a type I error occurs when a true null hypothesis is rejected, the simulated results ensure the type I error from a two-sided test at 0.05.

An interesting phenomenon was discovered in our simulation procedure. When comparing the estimators of $\widetilde{VUS}$ and $\widehat{VUS}$ from a single simulated dataset (data not shown here), the discrepancies between the two were more apparent under moderate association or non-informative association than under strong association, even though their means were always very close to each other across different associations. For example, the discrepancy between the two estimates from 1000 individual datasets that we simulated with 30% censoring could be as high as 26% in moderate association at $t_0 = 0.3$, and the maximum discrepancy between the two estimates was 18% for the no association case at $t_0 = 0.3$. In contrast, the discrepancy was only as much as 9% under the strong association at $t_0 = 0.05$.

In addition, the resulting estimators from both proposed methods demonstrated satisfactory behaviors in the simulation results in term of unbiasness and nominal level of coverage rates when $n = 150$ in Table 6.2. The standard errors from simulations of sample size of 150 were approximately $\sqrt{2}$ times of those from sample size of 300, indicating that the standard deviation goes to 0 as $n \to \infty$. Also, simulation results with tied scores in the biomarker (Table 6.3) were satisfactorily similar to those without tied scores. The proposed estimation using either approach is robust against tied scores in a biomarker if the tied biomarker is not systematic in favor of concordance or discordance.

Table 6.1: Simulation results of the two VUS estimators for the untied scores in biomarker with sample size n=300 (1000 replications)

			$\widetilde{VUS}$				$\widehat{VUS}$			
t_0	Cen rate	True AUC	$B_{\widetilde{VUS}}$	$ESE_{\widetilde{VUS}}$	$ASE_{\widetilde{VUS}}$	$CP_{\widetilde{VUS}}$	$B_{\widehat{VUS}}$	$ESE_{\widehat{VUS}}$	$BSE_{\widehat{VUS}}$	$CP_{\widehat{VUS}}$
Strong predicative power										
0.01	30%	0.542	0.002	0.032	0.034	95.4%	0.004	0.034	0.036	96.3%
0.05	30%	0.790	0.002	0.037	0.037	95.0%	0.004	0.039	0.038	95.0%
0.20	30%	0.760	0.002	0.037	0.040	95.4%	-0.010	0.040	0.042	95.5%
0.01	15%	0.542	0.000	0.030	0.034	96.2%	0.004	0.031	0.034	96.1%
0.05	15%	0.790	0.001	0.037	0.036	94.6%	0.001	0.038	0.037	94.5%
0.20	15%	0.760	0.000	0.032	0.035	95.8%	-0.007	0.033	0.037	96.4%
Moderate predictive power										
0.05	30%	0.296	0.001	0.038	0.038	94.8%	0.007	0.039	0.039	95.5%
0.20	30%	0.355	0.001	0.036	0.037	95.6%	0.005	0.037	0.039	96.2%
0.30	30%	0.382	0.001	0.046	0.044	93.0%	-0.001	0.038	0.039	95.7%
0.05	15%	0.296	0.000	0.038	0.037	94.2%	0.006	0.039	0.039	95.4%
0.20	15%	0.355	0.002	0.035	0.034	94.4%	0.004	0.035	0.035	94.7%
0.30	15%	0.382	0.001	0.036	0.038	96.2%	0.002	0.037	0.039	95.5%
No predictive power										
0.05	30%	0.167	0.000	0.032	0.030	93.8%	0.006	0.032	0.033	95.4%
0.20	30%	0.167	0.003	0.029	0.027	95.0%	0.003	0.028	0.029	95.5%
0.30	30%	0.167	0.001	0.038	0.036	94.2%	0.004	0.035	0.033	94.1%
0.05	15%	0.167	0.000	0.032	0.030	94.6%	0.003	0.032	0.033	95.7%
0.20	15%	0.167	-0.001	0.024	0.024	95.2%	0.005	0.025	0.027	95.4%
0.30	15%	0.167	0.000	0.031	0.029	94.6%	0.001	0.031	0.030	94.3%

Table 6.2: Simulation results of the two VUS estimators for the untied scores in biomarker with sample size n=150 (1000 replications)

			$\widetilde{VUS}$				$\widehat{VUS}$			
t_0	Cen rate	True AUC	$B_{\widetilde{VUS}}$	$ESE_{\widetilde{VUS}}$	$ASE_{\widetilde{VUS}}$	$CP_{\widetilde{VUS}}$	$B_{\widehat{VUS}}$	$ESE_{\widehat{VUS}}$	$BSE_{\widehat{VUS}}$	$CP_{\widehat{VUS}}$
Strong predicative power										
0.01	30%	0.542	0.002	0.046	0.050	96.2%	0.000	0.047	0.048	95.1%
0.05	30%	0.790	0.005	0.052	0.052	94.4%	-0.005	0.058	0.055	94.8%
0.20	30%	0.760	0.000	0.055	0.058	96.0%	-0.018	0.062	0.060	94.3%
0.01	15%	0.542	0.003	0.044	0.049	96.2%	0.001	0.045	0.045	94.6%
0.05	15%	0.790	0.001	0.049	0.050	95.2%	-0.003	0.052	0.050	94.3%
0.20	15%	0.760	-0.001	0.046	0.051	96.0%	-0.011	0.047	0.049	96.1%
Moderate predictive power										
0.05	30%	0.296	0.001	0.055	0.055	94.6%	0.001	0.056	0.057	94.6%
0.20	30%	0.355	0.003	0.055	0.053	93.8%	0.001	0.057	0.055	95.7%
0.30	30%	0.382	0.003	0.063	0.062	94.8%	-0.002	0.067	0.063	94.3%
0.05	15%	0.296	0.002	0.053	0.054	95.6%	0.003	0.053	0.051	94.5%
0.20	15%	0.355	0.001	0.048	0.049	94.8%	0.005	0.049	0.051	95.9%
0.30	15%	0.382	-0.005	0.053	0.054	94.6%	-0.004	0.054	0.056	95.6%
No predictive power										
0.05	30%	0.167	-0.001	0.046	0.043	93.8%	0.002	0.046	0.047	95.1%
0.20	30%	0.167	0.000	0.039	0.038	94.6%	0.001	0.039	0.038	95.3%
0.30	30%	0.167	0.001	0.051	0.051	95.2%	-0.001	0.047	0.048	94.5%
0.05	15%	0.167	0.001	0.042	0.043	95.4%	0.005	0.042	0.044	95.4%
0.20	15%	0.167	-0.001	0.036	0.036	95.6%	0.001	0.036	0.037	95.1%
0.30	15%	0.167	-0.002	0.044	0.042	94.6%	0.000	0.042	0.041	95.1%

Table 6.3: Simulation results of the two VUS estimators for the tied scores in biomarker with sample size n=300 (1000 replications)

			$\widetilde{VUS}$				$\widehat{VUS}$			
t_0	Cen rate	True AUC	$B_{\widetilde{VUS}}$	$ESE_{\widetilde{VUS}}$	$ASE_{\widetilde{VUS}}$	$CP_{\widetilde{VUS}}$	$B_{\widehat{VUS}}$	$ESE_{\widehat{VUS}}$	$BSE_{\widehat{VUS}}$	$CP_{\widehat{VUS}}$
Strong predicative power										
0.01	30%	0.542	-0.001	0.033	0.036	96.6%	-0.005	0.033	0.037	96.6%
0.05	30%	0.790	-0.001	0.038	0.037	93.6%	-0.006	0.041	0.040	94.3%
0.20	30%	0.760	0.000	0.034	0.038	96.2%	-0.016	0.037	0.039	95.9%
0.01	15%	0.542	0.002	0.030	0.034	97.1%	-0.003	0.032	0.037	96.7%
0.05	15%	0.790	0.003	0.035	0.035	95.2%	0.003	0.037	0.036	94.1%
0.20	15%	0.760	0.001	0.030	0.035	97.6%	-0.011	0.034	0.037	96.1%
Moderate predictive power										
0.05	30%	0.296	-0.001	0.038	0.039	95.7%	0.002	0.038	0.038	95.1%
0.20	30%	0.355	0.000	0.037	0.037	95.0%	0.002	0.039	0.038	94.1%
0.30	30%	0.382	0.002	0.044	0.044	95.1%	-0.001	0.046	0.045	94.1%
0.05	15%	0.296	0.001	0.037	0.037	94.1%	0.003	0.038	0.038	95.2%
0.20	15%	0.355	0.000	0.033	0.034	96.2%	0.001	0.035	0.034	94.7%
0.30	15%	0.382	0.001	0.038	0.038	94.2%	-0.002	0.040	0.040	95.3%
No predictive power										
0.05	30%	0.167	0.003	0.031	0.030	93.8%	0.006	0.031	0..031	94.5%
0.20	30%	0.167	0.000	0.027	0.027	94.5%	0.003	0.027	0.027	94.7%
0.30	30%	0.167	-0.003	0.033	0.034	94.6%	0.001	0.034	0.033	94.1%
0.05	15%	0.167	-0.001	0.030	0.029	94.1%	0.004	0.030	0.030	94.7%
0.20	15%	0.167	-0.001	0.024	0.024	94.6%	0.003	0.025	0.025	94.9%
0.30	15%	0.167	-0.001	0.029	0.029	94.0%	0.000	0.030	0.029	94.7%

6.3 DISCRIMINATIVE ABILITY OF VUS CORRESPONDING TO AUC

To address the discriminative ability of VUS, Table 6.4 presents the accuracy of a diagnostic test using the biomarker Y to the competing events with five predictive associations. We carry over the three scenarios in aforementioned simulation design, along with two additional predictive associations representing the predictive power from very strong, strong, moderate, weak, to none. The simulated datasets are generated with the parameters β_d and γ_d, for $d = 1, 2, 3$ which determine the underlying association between the cause-1 event, the cause-2 event and the biomarker Y, as well as t_0s which are chosen to classify the disease statuses without censoring for sample size 300. We reported $\widetilde{VUS}$, as well as corresponding $AUC_{1,3}$ (by comparing the cause-1 cases to the healthy controls), $AUC_{2,3}$ (by comparing the cause-2 cases to the healthy controls), and $AUC_{1,2}$ (by comparing the cause-1 cases to the cause-2 cases). All of the estimation of AUC were derived using the R package "pROC" due to the simulated data without censoring. From those reported $\widetilde{VUS}$, and corresponding AUCs, VUS, ranging from 0.167-1, is a metric reflecting the summary accuracy of the three AUC.

Table 6.4: Discriminative ability of VUS associated with AUC

Predictive power	Parameter	t_0	$\widetilde{VUS}$	$AUC_{1,3}$	$AUC_{1,2}$	$AUC_{1,2}$
Very strong	$\beta_1 = 50.0$; $\beta_2 = 3.0$; $\beta_3 = 1.0$; $\gamma_1 = 3.0$; $\gamma_2 = 50.0$; $\gamma_2 = 1.0$	0.05	0.874	0.971	0.935	0.936
Strong	$\beta_1 = 22.5$; $\beta_2 = 3.5$; $\beta_3 = 0.4$; $\gamma_1 = 3.5$; $\gamma_2 = 3.0$; $\gamma_2 = 1.8$	0.30	0.518	0.927	0.723	0.779
Moderate	$\beta_1 = 8.0$; $\beta_2 = 3.5$; $\beta_3 = 1.0$; $\gamma_1 = 3.5$; $\gamma_2 = 3.0$; $\gamma_2 = 2.0$	0.30	0.418	0.851	0.729	0.657
Weak	$\beta_1 = 4.5$; $\beta_2 = 3.2$; $\beta_3 = 1.5$; $\gamma_1 = 3.5$; $\gamma_2 = 3.0$; $\gamma_2 = 1.8$	0.30	0.316	0.76	0.723	0.544
None	$\beta_1 = 3.0$; $\beta_2 = 3.0$; $\beta_3 = 3.0$; $\gamma_1 = 3.0$; $\gamma_2 = 3.0$; $\gamma_2 = 3.0$	0.30	0.173	0.513	0.507	0.517

7.0 REAL STUDY EXAMPLES

7.1 THE MYHAT DATA

We applied our proposed methods to the cohort data of the Monongahela-Youghiogheny Healthy Aging Team (MYHAT). A random community sample with normal cognitive function to mild cognitive impairment was recruited beginning in year 2006, stratified by three age groups 65-74, 75-84, and 85+ years. The participants were followed prospectively for up to 9 annual visits to assess their cognitive decline and onset of dementia. The participants cognitive status was evaluated at baseline and the follow-up visits using the Clinical Dementia Rating (CDR). We classified $CDR \geq 0.5$ as the occurrence of cognitive impairment or dementia, and death as the other competing event. Death was treated as a worse medical condition. There were 1982 participants who met the study eligibility criteria and completed their cognitive test battery in follow-ups. For homogeneity, we removed the participants who were measured with $CDR \geq 0.5$ at baseline. Also, we excluded one deceased participant whose deceased date was prior to its baseline date, and our final analytical sample included 1412 participants.

We are interested in testing whether the five cognitive scores targeting in domains of attention, executive function, language, memory and visuospatial at baseline, can predict subsequent cognitive impairment and death within a 5 or 7 years. The cognitive scores were standardized by subtracting the age and sex adjusted population means and dividing by their standard deviations. According to the age-stratified design, our analysis was conducted for the three age subgroups, where we renamed the three age groups 65-74, 75-84, and 85+ years as young, middle and old age groups respectively. The frequencies of observed numbers of deaths, cognitive impairment, event-free survivors and censoring participants at

the selected t_0 of 5 years were reported in Table 7.1. Note that censored participants were used for estimation of weights in the IPCW method, who were not accounted for calculating percentages within the age groups.

We compared the distributions of IPCW-adjusted cognitive testing scores to their unweighted scores for young, middle and old age groups by the disease status of death, cognitive impairment, and cognitive-normal survivors at $t_0 = 5$ years in Tables 7.2, 7.3, 7.4. According to space limitation, we put Atten(tion), Execu(tive function), Langu(age), Memor(y), and Visuo(spatial) representing the five cognitive testing domains. From either methods, The distributions of those five cognitive tests didn't exhibit much discrimination trends across the three disease statuses using either methods, though IPCW-adjusted distributions were more directly related to the $\widetilde{VUS}$ in handling missingness due to right censoring. Moreover, we presented the estimated density plots for the five domain scores for those who experienced cognitive impairment (green dash lines), and those who died (red solid lines), as well as those who were alive and cognitively normal (blue dotted lines) after 5 years of follow up, in the MYHAT young, middle, old age groups in Figures 7.1, 7.2, and 7.3 respectively. In accounting for missingness due to right censoring before 5 years, we adopted the IPCW method by weighting domain scores for each disease status as in equation (4.1.1), and then applied the kernel smoothing method to generate the density plots using the R package "sm." In general, the three density curves were in general close to each other in young and middle age groups, indicating poor separability of the disease progression in five years based on those domain scores at baseline. However in the old age group, the three density curves looked more discriminatory than the two younger groups, although they still displayed significant overlaps in terms of areas under those three curves.

The VUSs from the two approaches, as well as two AUC estimates for each of the competing events (death or cognitive impairment) as compared to the event-free survivors were computed in Table 7.5, where the AUC was derived using the R package "timeROC". In general, the 5 cognitive test scores were shown to have poor predictive power in predicting cognitive impairment and death since the estimators from both methods were relatively low, which were considerably consistent with the results of AUCs. Along with what we have observed in the three Figures, we conclude that the five domain scores at baseline exhibit

poor predictive power in discriminating the sequence of the competing events. Nevertheless, VUSs allow the measurement of the global concordance between the continuous cognitive test scores and the sequence of cognitive impairment and death.

Moreover, the estimated VUSs from the two approaches are reasonably close and most discrepancies are less than one standard deviation. This is consistent with what we have discovered in simulations that the two estimators differ more in weaker association. We repeated the analyses at $t_0 = 7$ years and observed similar results. Interestingly, we observe stronger associations between baseline domain scores and disease progression in older groups at both time points, despite that the predictive power is generally weak. For example, estimated $\widetilde{VUS}$ for baseline executive function showed increasing predictive power with age as 0.176, 0.253, 0.303 for aged 65-74, 75-84, and 85+ years in the 5 years window.

Table 7.1: Frequencies of death, cognitive impairment, survivor and censored participants in each of the three age subgroups in MYHAT study ($t_0 = 5$ years)

Age group	Death	Dementia	Survivor	Censored
Young (n=539)	40 (9.4%)	37 (8.7%)	349 (81.9%)	113
Middle (n=654)	77 (15.3%)	131 (25.9%)	297 (58.8%)	149
Old (n=219)	64 (34.6%)	60 (32.4%)	61 (33.0%)	34

Table 7.2: Unweighted vs. IPCW-weighted distributions of cognitive testing scores in young age group by event status ($t_0 = 5$ years)

	Unweighted Distribution of Cog. Score					IPCW-weighted Distribution of Cog. Score				
Cognitive Test	Min.	25th Pctl	Median	75th Pctl	Max.	Min.	25th Pctl	Median	75th Pctl	Max.
Event Status = Death										
Atten	-2.245	-0.135	0.245	0.682	1.747	-2.599	-0.151	0.259	0.778	2.212
Execu	-1.443	-0.093	0.212	0.626	1.486	-1.562	-0.108	0.265	0.743	1.733
Langu	-1.531	0.131	0.524	0.752	1.494	-1.961	0.158	0.624	0.826	1.734
Memor	-1.044	-0.027	0.533	0.902	1.313	-1.344	-0.025	0.628	0.994	1.604
Visuo	-1.752	-0.262	0.430	1.121	2.505	-1.894	-0.241	0.485	1.269	3.201
Event Status = Cognitive Impairment										
Atten	-1.731	-0.186	0.361	0.783	1.747	-1.785	-0.229	0.424	0.869	1.908
Execu	-1.066	-0.244	0.202	0.777	1.703	-1.246	-0.289	0.212	0.793	1.991
Langu	-1.383	-0.232	0.328	0.734	1.204	-1.617	-0.275	0.383	0.769	1.271
Memor	-1.435	-0.434	0.012	0.591	1.449	-1.677	-0.459	0.013	0.603	1.601
Visuo	-1.965	-0.900	0.004	1.015	2.718	-2.412	-0.032	0.005	1.083	3.179
Event Status = Survivors										
Atten	-1.295	-0.010	0.433	0.894	2.780	-1.673	-0.013	0.559	1.155	3.591
Execu	-1.230	0.119	0.498	0.889	2.912	-1.589	0.156	0.643	1.141	3.761
Langu	-3.132	0.231	0.555	0.839	1.770	-0.405	0.303	0.716	1.083	2.286
Memor	-1.475	0.229	0.615	0.979	2.015	-1.905	0.297	0.794	1.262	2.603
Visuo	-2.603	-0.156	0.376	1.121	3.462	-3.347	-0.201	0.484	1.442	4.452

Table 7.3: Unweighted vs. IPCW-weighted distributions of cognitive testing scores in middle age group by event status ($t_0 = 5$ years)

	Unweighted Distribution of Cog. Score					IPCW-weighted Distribution of Cog. Score				
Cognitive Test	Min.	25th Pctl	Median	75th Pctl	Max.	Min.	25th Pctl	Median	75th Pctl	Max.
Event Status = Death										
Atten	-1.386	-0.534	-0.069	0.361	1.138	-1.504	-0.611	-0.081	0.431	1.425
Execu	-2.686	-0.566	-0.081	0.298	1.275	-2.864	-0.648	-0.094	0.374	1.281
Langu	-2.207	-0.209	0.086	0.486	1.176	-2.461	-0.227	0.106	0.609	1.315
Memor	-1.633	-0.439	-0.042	0.411	1.303	-1.889	-0.519	-0.053	0.446	1.481
Visuo	-1.752	-0.688	-0.049	0.376	1.334	-2.359	0.779	-0.063	0.457	1.542
Event Status = Cognitive Impairment										
Atten	-1.766	-0.572	-0.050	0.470	1.648	-2.097	-0.637	-0.053	0.567	2.087
Execu	-1.689	-0.551	-0.106	0.254	2.697	-1.847	-0.641	-0.115	0.272	3.416
Langu	-1.850	-0.423	0.055	0.348	1.538	-1.918	-0.439	0.059	0.429	1.948
Memor	-1.976	-0.605	-0.232	0.160	1.343	-2.034	-0.683	-0.251	0.191	1.831
Visuo	-2.071	-0.794	-0.049	0.376	2.611	-2.285	-0.914	-0.065	0.408	3.578
Event Status = Survivors										
Atten	-1.618	-0.397	0.029	0.541	2.244	-2.251	-0.552	0.041	0.752	3.006
Execu	-2.151	-0.166	0.257	0.579	2.367	-2.988	-0.231	0.357	0.804	3.287
Langu	-1.727	-0.080	0.286	0.665	1.494	-2.398	-0.111	0.398	0.924	2.074
Memor	-1.372	-0.058	0.274	0.664	1.955	-1.908	-0.081	0.382	0.923	2.719
Visuo	-1.752	-0.368	0.164	0.696	3.037	-2.426	-0.437	0.227	0.964	4.206

Table 7.4: Unweighted vs. IPCW-weighted distributions of cognitive testing scores in old age group by event status ($t_0 = 5$ years)

	Unweighted Distribution of Cog. Score					IPCW-weighted Distribution of Cog. Score				
Cognitive Test	Min.	25th Pctl	Median	75th Pctl	Max.	Min.	25th Pctl	Median	75th Pctl	Max.
Event Status = Death										
Atten	-1.698	-1.020	-0.565	0.066	1.325	-1.945	-1.195	-0.619	0.045	1.452
Execu	-2.783	-0.942	-0.558	-0.095	1.042	-3.487	-1.066	-0.647	-0.115	1.088
Langu	-2.924	-0.845	-0.305	0.022	1.052	-3.531	-0.889	-0.354	0.021	1.271
Memor	-1.463	-0.936	-0.513	-0.147	1.433	-1.671	-0.999	-0.584	-0.158	1.795
Visuo	-1.752	-0.900	-0.475	0.057	1.441	-2.121	-0.987	-0.493	0.058	1.743
Event Status = Cognitive Impairment										
Atten	-1.982	-0.572	-0.039	0.509	1.434	-2.335	-0.944	-0.511	0.255	1.689
Execu	-1.346	-0.445	-0.140	0.346	1.862	-2.233	-0.721	-0.496	0.041	1.972
Langu	-2.715	-0.431	-0.037	0.279	1.494	-2.934	-1.023	-0.428	0.029	1.067
Memor	-1.446	-0.415	-0.064	0.560	1.333	-1.997	-1.185	-0.628	0.225	1.434
Visuo	-2.177	-0.475	-0.049	0.376	1.547	-2.441	-1.362	-0.916	-0.147	1.573
Event Status = Survivors										
Atten	-1.982	-0.572	-0.039	0.509	1.434	-2.618	-0.756	-0.052	0.667	1.894
Execu	-1.346	-0.445	-0.140	0.346	1.862	-1.773	-0.586	-0.185	0.455	2.453
Langu	-2.715	-0.431	-0.037	0.279	1.494	-3.576	-0.568	-0.049	0.368	1.968
Memor	-1.446	-0.415	-0.064	0.560	1.333	-1.905	-0.546	-0.084	0.738	1.756
Visuo	-2.177	-0.475	-0.049	0.376	1.547	-2.897	-0.632	-0.065	0.501	2.058

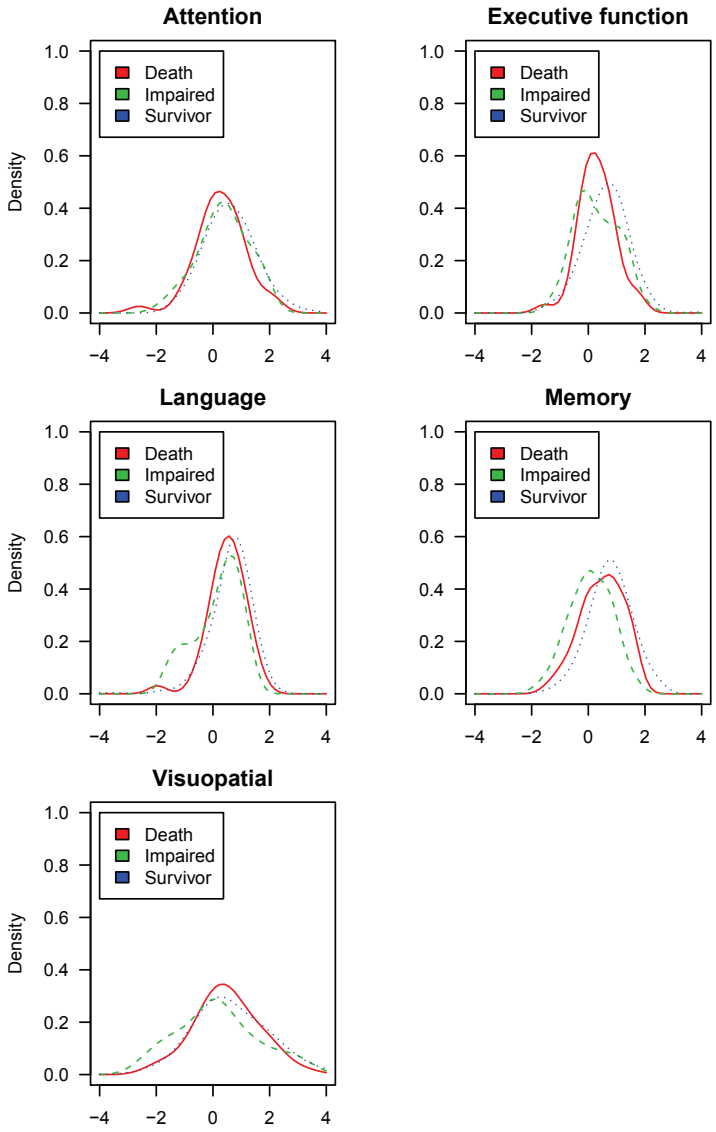

Figure 7.1: IPCW-adjusted distributions of five cognitive test scores in Young Age Group by disease status in MAHAT study

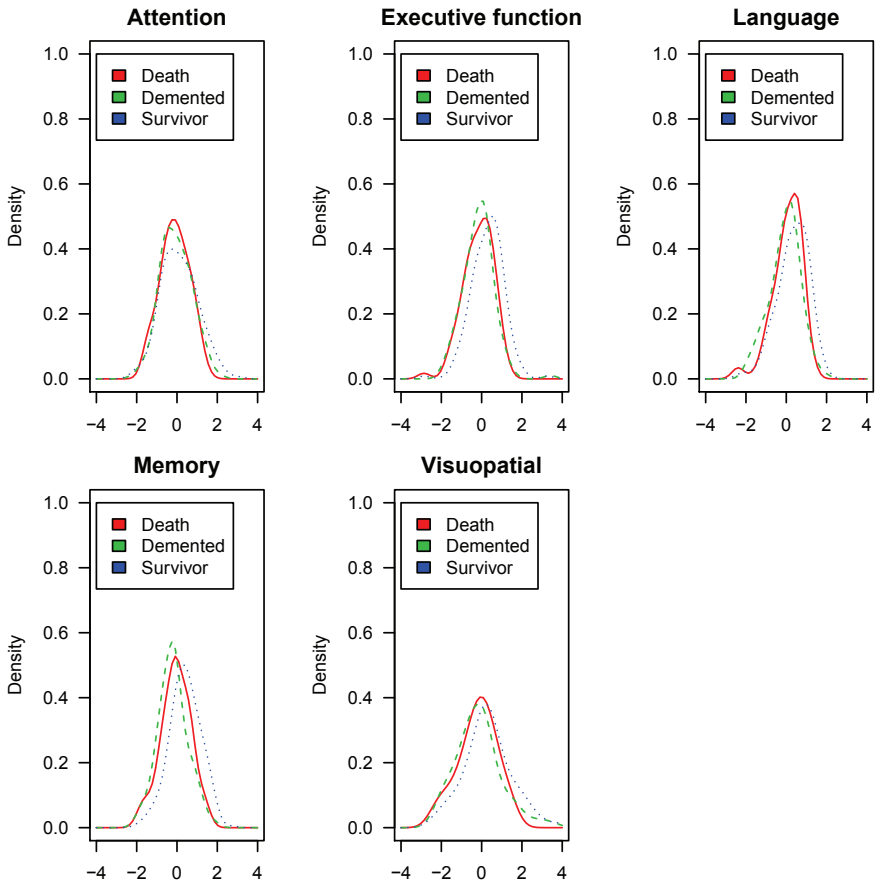

Figure 7.2: IPCW-adjusted distributions of five cognitive test scores in Middle Age Group by disease status in MAHAT study

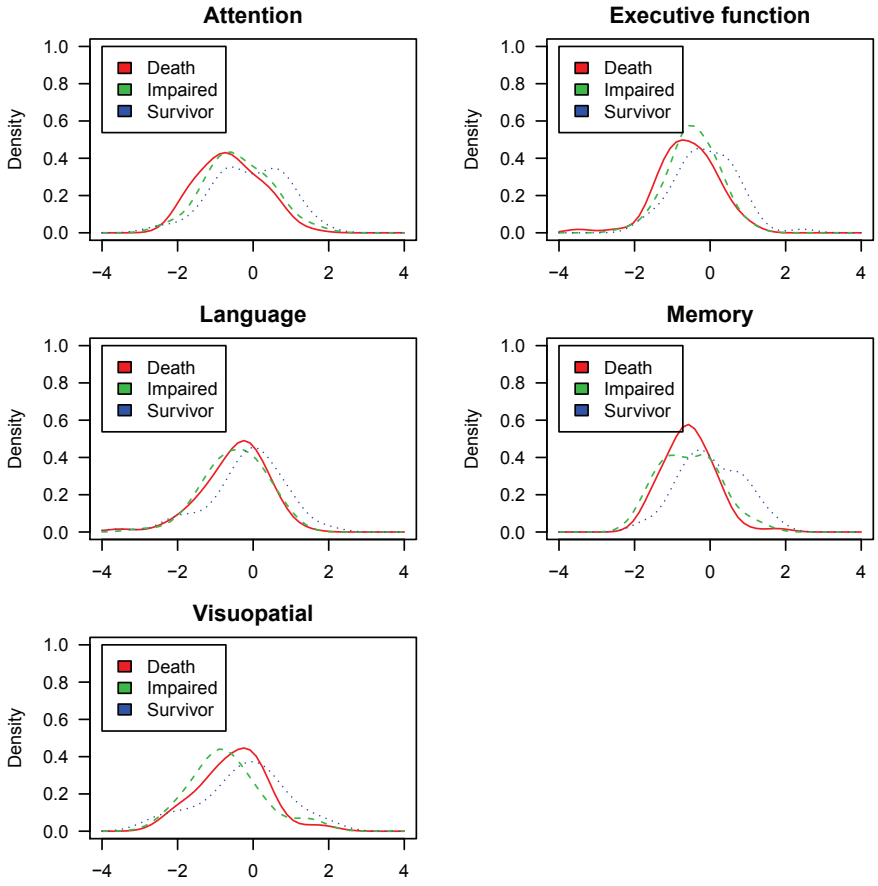

Figure 7.3: IPCW-adjusted distributions of five cognitive test scores in Old Age Group by disease status in MAHAT study

Table 7.5: VUS and AUC of five cognitive testing scores at $t_0 = 5$ years

age group	Cognitve test	$\widetilde{VUS}$ ($\hat{\sigma}_{\widetilde{VUS}}$)	$\widetilde{VUS}$ ($\hat{\sigma}_{\widetilde{VUS}}$)	AUC_1 ($\hat{\sigma}_{AUC_1}$)	AUC_2 ($\hat{\sigma}_{AUC_2}$)
Young	Attention	0.190 (0.030)	0.170 (0.032)	0.581 (0.048)	0.538 (0.054)
	Executive	0.176 (0.030)	0.151 (0.027)	0.623 (0.044)	0.594 (0.056)
	Language	0.168 (0.030)	0.160 (0.033)	0.547 (0.049)	0.636 (0.051)
	Memory	0.184 (0.036)	0.181 (0.037)	0.572 (0.051)	0.722 (0.046)
	Visuosptial	0.139 (0.026)	0.146 (0.029)	0.510 (0.048)	0.591 (0.058)
Middle	Attention	0.194 (0.022)	0.195 (0.021)	0.571 (0.036)	0.556 (0.031)
	Executive	0.253 (0.027)	0.233 (0.026)	0.659 (0.035)	0.671 (0.029)
	Language	0.199 (0.024)	0.175 (0.022)	0.589 (0.035)	0.637 (0.029)
	Memory	0.234 (0.028)	0.221 (0.026)	0.656 (0.036)	0.750 (0.028)
	Visuosptial	0.216 (0.024)	0.216 (0.031)	0.622 (0.037)	0.623 (0.032)
Old	Attention	0.276 (0.037)	0.272 (0.033)	0.684 (0.048)	0.619 (0.051)
	Executive	0.303 (0.039)	0.273 (0.038)	0.690 (0.047)	0.648 (0.050)
	Language	0.228 (0.034)	0.225 (0.034)	0.633 (0.050)	0.633 (0.051)
	Memory	0.255 (0.036)	0.248 (0.037)	0.730 (0.045)	0.695 (0.048)
	Visuosptial	0.209 (0.033)	0.247 (0.038)	0.612 (0.053)	0.665 (0.053)

1. AUC_1 compares deaths to healthy controls.

2. AUC_2 compares cognitive impairment to healthy controls.

7.2 THE PBC DATA

We next applied our methods to a well-known public dataset from the Mayo Clinical trial of Primary Biliary Cirrhosis (PBC) of the liver. The trial was conducted between 1974 and 1984, and the analysis included 418 patients who met eligibility criteria with complete data. Bilirubin is a prognostic biomarker commonly applied in PBC since cirrhosis causes high bilirubin levels. In the PBC data, bilirubin was heavily tied especially at levels below 2 mg/dL. Here, we reversed bilirubin values using a reciprocal function in order to be consistent with the direction in our definitions of VUS. The objective of the analysis was to determine the time window by which bilirubin could reasonably predict liver transplant (cause-2) and death (cause-1) events. Since liver transplants typically took place after 500 days of follow-up while death occurred from day 41, we picked 5 time points as $t_0 = 1000, 1500, 2000, 2500, 3000$ days.

The frequency of observed numbers of deaths, liver transplantation, event-free survivors and censoring patients at the different selected t_0 were reported in Table 7.6. As before, the censored participants were used for estimation of weights in the IPCW method, and we didn't account them for percentages. We provided the unweighted and IPCW-adjusted distributions of bilirubin across all selected times in Table 7.7. Based on the estimated VUSs and AUCs (in Table 7.8), bilirubin presented an overall predictive power on both events across those selected times, with highest discriminative power during 4-5.5 years. Consistently, the three IPCW-adjusted density curves of the reciprocal bilirubin values for those who died, had liver transplant, or survived without transplant were most separated at $t_0 = 1500$ days; see Figure 7.4 for details. As expected, the estimates were closer than the ones in MYHAT study, as the associations between bilirubin and competing events were generally stronger.

Table 7.6: Frequencies of death, liver transplant, survivor and censored patients at different t_0 in PBC study (n=418).

t_0 (days)	Death	Liver transplant	Survivor	Censored
1000	76 (18.5%)	7 (1.7%)	327 (79.8%)	8
1500	104 (29.1%)	14 (3.9%)	240 (67.0%)	60
2000	118 (37.7 %)	17 (5.4%)	178 (56.9%)	105
2500	134 (47.8%)	23 (8.2%)	123 (43.9%)	138
3000	143 (59.1%)	23 (9.5%)	76 (31.4%)	176

Table 7.7: Unweighted vs. IPCW-weighted distributions of bilirubin (transformed with a reciprocal function) at different t_0 by event status

	Unweighted Distribution of Cognitive Score					IPCW-weighted Distribution of Cognitive Score				
t_0 (days)	Min.	25th Pctl	Median	75th Pctl	Max.	Min.	25th Pctl	Median	75th Pctl	Max.
Event Status = Death										
1000	0.036	0.086	0.211	0.389	1.667	0.036	0.086	0.211	0.395	1.667
1500	0.036	0.089	0.211	0.421	1.667	0.036	0.089	0.223	0.421	1.753
2000	0.036	0.107	0.233	0.450	1.667	0.036	0.113	0.248	0.491	1.753
2500	0.036	0.128	0.271	0.548	3.333	0.036	0.138	0.298	0.638	6.385
3000	0..036	0.137	0.286	0.590	3.333	0.036	0.141	0.313	0.769	6.385
Event Status = Liver Transplant										
1000	0.164	0.205	0.313	0.369	0.833	0.164	0.205	0.315	0.372	0.833
1500	0.115	0.242	0.369	0.616	1.250	0.119	0.245	0.372	0.739	1.281
2000	0.115	0.286	0.323	0.588	1.250	0.119	0.296	0.364	0.676	1.281
2500	0.056	0.257	0.313	0.607	2.000	0.095	0.306	0.421	0.638	3.386
3000	0.056	0.257	0.313	0.607	2.000	0.095	0.306	0.421	0.718	3.386
Event Status = Survivors										
1000	0.051	0.455	0.909	1.429	3.333	0.051	0.467	0.933	1.467	3.423
1500	0.056	0.526	1.000	1.429	3.333	0.069	0.651	1.237	1.767	4.124
2000	0.056	0.635	1.111	1.667	3.333	0.086	0.981	1.715	2.572	5.144
2500	0.069	0.691	1.111	1.429	3.333	0.135	1.345	2.164	2.782	6.491
3000	0.112	0.714	1.111	1.429	3.333	0.325	2.021	3.189	4.101	9.566

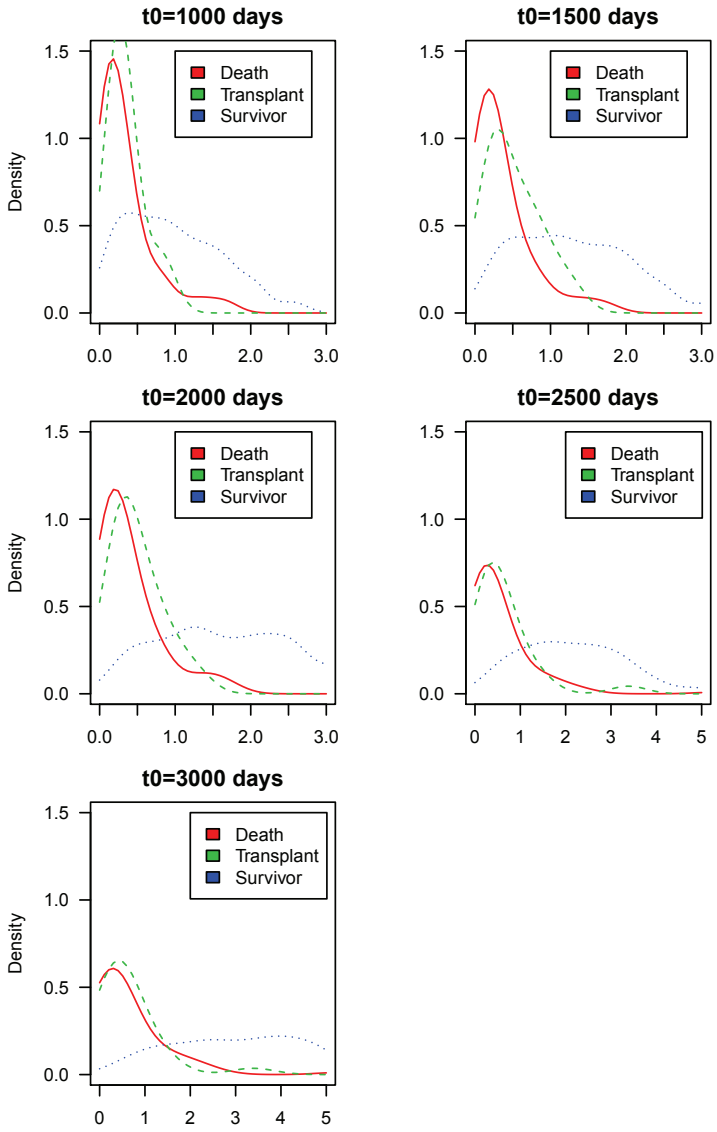

Figure 7.4: IPCW-adjusted distributions of bilirubin (transformed with a reciprocal function) by disease status at different t_0 in PBC Study

Table 7.8: VUS and AUC of bilirubin at t_0= 1000, 1500, 2000, 2500, 3000 days.

t_0 (days)	$\widetilde{VUS}$ ($\hat{\sigma}_{\widetilde{VUS}}$)	$\widehat{VUS}$ ($\hat{\sigma}_{\widehat{VUS}}$)	AUC_1 ($\hat{\sigma}_{AUC_1}$)	AUC_2 ($\hat{\sigma}_{AUC_2}$)
1000	0.486 (0.045)	0.486 (0.047)	0.823 (0.027)	0.799 (0.050)
1500	0.510 (0.039)	0.499 (0.042)	0.856 (0.022)	0.780 (0.049)
2000	0.513 (0.039)	0.502 (0.038)	0.864 (0.022)	0.827 (0.040)
2500	0.416 (0.046)	0.411 (0.047)	0.820 (0.028)	0.816 (0.560)
3000	0.390 (0.046)	0.384 (0.047)	0.805 (0.031)	0.822 (0.058)

1. AUC_1 compares deaths to healthy controls.

2. AUC_2 compares liver transplant to healthy controls.

8.0 CONCLUSIONS

In this dissertation we introduced the concept of the ROC surface and the associated volume under the ROC surface (VUS) in measuring diagnostic accuracy of a continuous biomarker to competing risks outcomes, where the samples were also subject to random censoring. Based on two alternative aspects of viewing the VUS as the volume under the ROC surface that is spanned by correct classification probabilities or the concordance between a continuous biomarker and a sequence of competing events, we proposed two nonparametric estimators of the VUS, and established their asymptotic properties and illustrated their performance through simulations. We showed that even in the presence of tied biomarker values, the two approaches provide consistent estimates of VUS for censored data of varying proportions. Moreover, we applied the two VUS estimators to two study data sets, and the separate AUCs for each event comparing the healthy survivors showed consistent results with the estimated VUSs. Thus, the VUS provides a complementary global discriminatory metric for assessing ordered events simultaneously.

Both estimators exhibit satisfactory properties given the results from various simulation scenarios and real data analyses, however $\widetilde{VUS}$ should outperform $\widehat{VUS}$ for the following reasons. First, the model-based standard errors are directly produced for $\widetilde{VUS}$ given the implemented R program, while the estimated standard errors of $\widehat{VUS}$ are derived from bootstrap. Second, $\widetilde{VUS}$ is more reliable when subjects in each disease status categories are substantially unbalanced. $\widehat{VUS}$ is estimated according to summation over the uniform squares with respect to approximation of the volume under ROC surface, which leads to the resulting $\widehat{VUS}$ lack of precision when there are much fewer subjects in the disease group. In contrast, $\widetilde{VUS}$ is more reliable to unbalance samples. Third, regarding the solutions of the both estimators to handle tied scores in a biomarker, $\widetilde{VUS}$ is modified with corresponding

weights targeting in each specific tied score scenarios. $\widehat{VUS}$, in general, leads to an unbiased estimator by assuming a random pattern of tied score, which causes overestimation and underestimation of $\widehat{VUS}$ washing off to each other. However, for a specfic sample, it exists a potential in terms of overestimation (or underestimation) of $\widehat{VUS}$.

One important task in clinical decision making is to identify those subjects who will develop a certain severity in their disease progression by a specific time window, thus providing targets for better treatment or prevention. The concept of the VUS allows projecting a patient's diagnostic biomarker onto three-stage disease progression. It is straightforward to extend the metrics to higher dimensional outcomes by introducing multi-way ROC analysis (Scurfield, 1996) and the hypervolume under the ROC manifold (HUM). The HUM can also be estimated as a U-type statistic, as well as by integrating CCPs from multiple dimensions. Large sample properties such as consistency and weak convergence still hold for the estimated HUM, and the discussions on handling tied scores in a biomarker can be easily carried over to HUM estimators.

Currently we included a single biomarker to predict a sequence of competing events. In MYHAT study, we evaluated discriminatory power of the five cognitive test scores on cognitive impairment and death stratified by age. It may be more informative in treating age as a covariate rather than a stratified factor. To achieve this goal, we can fit semiparametric regression models as in Zheng et.al (2012). Moreover, if the competing events are not sequential, the biomarker should not be a scaler, but rather a vector(Obuchowski, 2005). Subjects are assigned to the disease status category with the highest probability given the vector of risk factors, which is estimated through some regression models. The VUS can then be defined as the concordance between the assigned disease category and the true disease status. Therefore, the extension of the VUS to nominal competing outcomes is technically straight-forward. This will be a topic of future research.

BIBLIOGRAPHY

Blanche, P., J.-F. Dartigues, and H. Jacqmin-Gadda (2013a). Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Statistics in Medicine 32*(30), 5381–5397.

Blanche, P., J.-F. Dartigues, and H. Jacqmin-Gadda (2013b). Review and comparison of roc curve estimators for a time-dependent outcome with marker-dependent censoring. *Biometrical Journal 55*(5), 687–704.

Cheng, Y., J. P. Fine, and M. R. Kosorok (2007). Nonparametric analysis of bivariate competing risks data. *Journal of the American Statistical Association 102*, 1407–1416.

Cox, D. R. (1959). The analysis of exponentially distributed life-times with two types of failure. *Journal of the Royal Statistical Society, Series B 21*, 411–421.

Dabrowska, D. M. (1988). Kaplan-meier estimate on the plane. *The Annals of Statistics*, 1475–1489.

Dinse, G. E. (1985). An alternative to efron's redistribution-of-mass construction of the kaplanmeier estimator. *The American Statistician 39*(4), 299–300.

Dreiseitl, S., O. L, and M. Binder (2000). Comparing three-class diagnostic tests by three-way ROC analysis. *Medical Decision Making 20*(3), 323–331.

Efron, B. (1967). The two sample problem with censored data. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, Volume 4, pp. 831–853.

Faraggi, D. and B. Reiser (2002). Estimation of the area under the roc curve. *Statistics in Medicine 21*(20), 3093–3106.

Fine, J. P. and R. J. Gray (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association 94*, 496–509.

Gooley, T. A., W. Leisenring, J. Crowley, and B. E. Storer (1999). Estimation of failure probabilities in the presence of competing risks: New representations of old estimators. *Statistics in Medicine 18*, 695–706.

Heagerty, P., T. Lumley, and M. S. Pepe (2000). Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics 56*, 337–344.

Heagerty, P. and Y. Zheng (2005). Survival model predictive accuracy and ROC curves. *Biometrics 61*, 92–105.

Horton, B. (2016). Calculating auc:the area under a roc curve. http://blog.revolutionanalytics.com/2016/11/calculating-auc.html.

Hung, H. and C.-T. Chiang (2010). Estimation methods for time-dependent auc models with survival data. *Canadian Journal of Statistics 38*(1), 8–26.

Kalbfleisch, J. D. and R. L. Prentice (2011). *The statistical analysis of failure time data*, Volume 360. John Wiley & Sons.

Lee, J. (1990). U-statistics: Theory and practice.

Li, J. and J. P. Fine (2008). ROC analysis with multiple classes and multiple tests: methodology and its application in microarray studies. *Biostatistics 9*(3), 566–576.

Li, J. and X.-H. Zhou (2009). Nonparametric and semiparametric estimation of the three way receiver operating characteristic surface. *Journal of Statistical Planning and Inference 139*(12), 4133–4142.

Li, R., Y. Cheng, Q. Chen, and J. Fine (2017). Quantile association for bivariate survival data. *Biometrics 73*(2), 506–516.

Lin, D. (1997). Non-parametric inference for cumulative incidence functions in competing risks studies. *Statistics in medicine 16*(8), 901–910.

Mossman, D. (1999). Three-way rocs. *Medical Decision Making 19*(1), 78–89.

Obuchowski, N. A. (2005). Estimating and comparing diagnostic tests accuracy when the gold standard is not binary. *Academic Radiology 12*(9), 1198–1204.

Peng, L. and Y. Huang (2008). Survival analysis with quantile regression models. *Journal of the American Statistical Association 103*(482), 637–649.

Pepe, M. S. (1991). Inference for events with dependent risks in multiple endpoint studies. *Journal of the American Statistical Association 86*(415), 770–778.

Prentice, R. L., J. D. Kalbfleisch, A. V. Peterson, N. Flournoy, V. T. Farewell, and N. E. Breslow (1978). The analysis of failure time data in the presence of competing risks. *Biometrics 12*, 737–751.

Saha, P. and P. J. Heagerty (2010). Time-dependent predictive accuracy in the presence of competing risks. *Biometrics 66*, 999 – 1011.

Scheike, T. H., M. J. Zhang, and T. A. Gerds (2008). Predicting cumulative incidence probability by direct binomial regression. *Biometrika 95*(1), 205–220.

Scurfield, B. K. (1996, September). Multiple-event forced-choice tasks in the theory of signal detectability. *J. Math. Psychol. 40*(3), 253–269.

Shi, H., Y. Cheng, and J. Li (2014). Assessing diagnostic accuracy improvement for survival or competing-risk censored outcomes. *Canadian Journal of Statistics 42*(1), 109–125.

Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences 72*, 20–22.

Van der Laan, M. J. and J. M. Robins (2003). *Unified methods for censored longitudinal data and causality*. Springer Science & Business Media.

Van Der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics, 3. Cambridge University Press.

Wang, H. and Y. Cheng (2014). Piecewise cause-specific association analyses of multivariate untied or tied competing risks data. *The international journal of biostatistics*.

Wolbers, M., P. Blanche, M. T. Koller, J. C. M. Witteman, and T. A. Gerds (2014). Concordance for prognostic models with competing risks. *Biostatistics 15*(3), 526–539.

Wu, Y. and C. Chiang (2013). Optimal receiver operating characteristic manifolds. *Journal of Mathematical Psychology 57*(5), 237–248.

Zhang, M.-J. and J. Fine (2008). Summarizing differences in cumulative incidence functions. *Statistics in medicine 27*(24), 4939–4949.

Zheng, Y., T. Cai, Y. Jin, and Z. Feng (2012). Evaluating prognostic accuracy of biomarkers under competing risk. *Biometrics 68*, 388–396.